

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

MATHEUS LISARTE BEZERRA

**SISTEMA INTELIGENTE NÃO SUPERVISIONADO PARA CLASSIFICAÇÃO DE
IMAGÉTICA MOTORA POR MEIO DE SINAIS DE EEG**

CURITIBA

2026

MATHEUS LISARTE BEZERRA

**SISTEMA INTELIGENTE NÃO SUPERVISIONADO PARA CLASSIFICAÇÃO DE
IMAGÉTICA MOTORA POR MEIO DE SINAIS DE EEG**

**Unsupervised Intelligent System for Motor Imagery Classification Using EEG
Signals**

Trabalho de conclusão de curso de graduação
apresentado como requisito para obtenção do título
de Bacharel em Engenharia Elétrica do curso de
Engenharia Elétrica da Universidade Tecnológica
Federal do Paraná (UTFPR).
Orientador: Narco Afonso Ravazzoli Maciejewski.
Coorientador: Jefferson Tales Oliva

CURITIBA

2026



Esta licença permite remixe, adaptação e criação a partir do trabalho, para fins não comerciais, desde que sejam atribuídos créditos ao(s) autor(es). Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

MATHEUS LISARTE BEZERRA

**SISTEMA INTELIGENTE NÃO SUPERVISIONADO PARA CLASSIFICAÇÃO DE
IMAGÉTICA MOTORA POR MEIO DE SINAIS DE EEG**

Trabalho de conclusão de curso de graduação
apresentado como requisito para obtenção do título
de Bacharel em Engenharia Elétrica do curso de
Engenharia Elétrica da Universidade Tecnológica
Federal do Paraná (UTFPR).

Data de aprovação: 23/junho/2026

Narco Afonso Ravazzoli Maciejewski
Doutorado
Universidade Tecnológica Federal do Paraná

Glauber Gomes De Oliveira Brante
Doutorado
Universidade Tecnológica Federal do Paraná

Renata Coelho Borges
Doutorado
Universidade Tecnológica Federal do Paraná

CURITIBA

2026

Dedico este trabalho à minha falecida irmã, minha maior fonte de inspiração e símbolo de força.

AGRADECIMENTOS

Certamente estas palavras não serão suficientes para contemplar todas as pessoas que fizeram parte desta importante fase da minha vida. Portanto, desde já peço desculpas àquelas que não estão presentes entre essas linhas, mas podem estar certas de que fazem parte do meu pensamento e da minha gratidão.

Agradeço ao meu orientador, Prof. Dr. Narco Afonso Ravazzoli Maciejewski, por ter me guiado com sabedoria não apenas nos desafios acadêmicos, mas também nos pessoais. Mais do que um professor excepcional, revelou-se uma pessoa extraordinária, cuja presença tornou esta jornada mais leve e significativa.

Aos meus amigos, pelo companheirismo genuíno nos momentos em que a confiança escasseava. Saber que havia pessoas dispostas a caminhar ao lado, especialmente quando o caminho pesava, fez toda a diferença.

À minha mãe, pelo apoio incondicional em cada tropeço, em cada recomeço. Sua presença constante me ensinou que persistir é, acima de tudo, um ato de amor.

À minha irmã, que partiu cedo demais, mas deixou uma herança que nenhuma distância apaga: o exemplo da sua luta, da sua força e da sua coragem diante da vida. Você continua sendo minha inspiração todos os dias.

À Coordenação e ao grupo docente do Curso, pela cooperação e disponibilidade ao longo desta trajetória.

"Aquele que tem um porquê para viver
suporta quase qualquer como."
(Nietzsche; Hollingdale, 1990).

RESUMO

Interfaces cérebro-computador baseadas em imagética motora permitem comunicação e controle sem movimentos físicos, mas enfrentam o desafio da variabilidade de desempenho entre os usuários. Este trabalho investigou a estrutura de sinais de eletroencefalografia em tarefas de imagética motora por meio de técnicas de aprendizado não supervisionado, visando identificar se os padrões naturais de agrupamento dos dados correspondem às classes experimentais definidas (mão esquerda e mão direita) e elucidar fatores relacionados à heterogeneidade interindividual. Utilizou-se o banco de dados público de Cho et al. (2017), que contém sinais de 52 sujeitos saudáveis, registrados com 68 eletrodos a 512 Hz durante 100 a 120 ensaios por classe. Dois *pipelines* de processamento e inferência foram propostas e comparadas: uma de referência, baseada em configurações consolidadas na literatura, composta por filtragem passa-baixas em 30 Hz, normalização *z-score*, janela temporal de 0,5 a 2,5 segundos, extração de energia e entropia por transformada *wavelet* discreta e parâmetros de Hjorth, redução de dimensionalidade por análise de componentes principais e agrupamento por *k-means*; e uma de meta-aprendizagem, que explorou sistematicamente 75 combinações de filtros, janelas temporais e normalizações, com extração de características por transformada *wavelet* discreta, parâmetros de Hjorth e biblioteca de extração de séries temporais, seleção por correlação de Pearson e seleção de características baseada em correlação, e agrupamento por *k-means*. O *pipeline* de referência apresentou um índice médio de Rand ajustado de 0,0062, com desempenho próximo de zero na maioria dos sujeitos, evidenciando limitações na generalização à população estudada. O *pipeline* de meta-aprendizagem superou expressivamente o de referência, com a melhor combinação, composta por um filtro passa-baixas de 30 Hz, normalização min-máx e janela atual, alcançando um índice médio de Rand ajustado de 0,1454. A análise do impacto das etapas de pré-processamento revelou que o filtro é o fator de maior influência, com filtros de banda ampla superando os passa-faixa em aproximadamente uma ordem de grandeza, e que janelas temporais mais longas produzem representações mais discriminativas. A análise das características selecionadas evidenciou a dominância do canal C3 em todas as combinações avaliadas, com concentração de características espectrais na banda alfa/mu e medidas estatísticas de dispersão no domínio do tempo, resultado neurologicamente coerente com a natureza bilateral da tarefa e consistente com a literatura de interfaces cérebro-computador. Os resultados obtidos contribuem para a compreensão dos mecanismos neurais subjacentes à imagética motora à variabilidade intersujeito, fornecendo subsídios para o desenvolvimento de sistemas de interface cérebro-computador mais adaptativos e personalizados.

Palavras-chave: Imagética Motora; Eletroencefalograma; Sistema Não Supervisionado; *Wavelet*.

ABSTRACT

Motor imagery-based brain-computer interfaces enable communication and control without physical movement yet face the challenge of performance variability across users. This work investigated the structure of electroencephalography signals during motor imagery tasks through unsupervised learning techniques, aiming to identify whether the natural clustering patterns of the data correspond to the experimentally defined classes (left hand and right hand) and to elucidate factors related to interindividual heterogeneity. The public dataset of Cho et al. (2017) was used, comprising signals from 52 healthy subjects recorded with 68 electrodes at 512 Hz during 100 to 120 trials per class. Two processing and inference pipelines were proposed and compared: a reference pipeline, based on configurations established in the literature, composed of low-pass filtering at 30 Hz, z-score normalization, a temporal window from 0.5 to 2.5 seconds, energy and entropy extraction via discrete wavelet transform and Hjorth parameters, dimensionality reduction by principal component analysis, and k-means clustering; and a meta-learning pipeline, which systematically explored 75 combinations of filters, temporal windows, and normalization methods, with feature extraction via discrete wavelet transform, Hjorth parameters, and a time-series feature extraction library, pre-filtering based on Pearson correlation, correlation-based feature selection, and k-means clustering. The reference pipeline yielded a mean adjusted Rand index of 0.0062, with performance near zero for most subjects, evidencing limitations in generalization to the studied population. The meta-learning pipeline substantially outperformed the reference, with the best combination, composed of a 30 Hz low-pass filter, min-max normalization, and the current temporal window, achieving a mean adjusted Rand index of 0.1454. Analysis of the impact of the preprocessing stages revealed that filtering is the most influential factor, with broadband filters outperforming bandpass filters by approximately one order of magnitude, and that longer temporal windows produce more discriminative representations. Analysis of the selected features evidenced the dominance of channel C3 across all evaluated combinations, with a concentration of spectral features in the alpha/mu band and statistical dispersion measures in the time domain, a result neurologically consistent with the bilateral nature of the task and consistent with the brain-computer interface literature. The findings contribute to the understanding of the neural mechanisms underlying motor imagery and intersubject variability, providing insights for the development of more adaptive and personalized brain-computer interface systems.

Keywords: Motor imagery; Electroencephalogram; Unsupervised learning; Wavelet.

LISTA DE ILUSTRAÇÕES

| | |
|---|----|
| Figura 1 - Distribuição dos eletrodos no sistema 10-20. | 20 |
| Figura 2 - Distribuição dos eletrodos no sistema 10-10 com 64 canais..... | 21 |
| Figura 3 - Padrões de onda EEG..... | 23 |
| Figura 4 - Mapeamento do plano s no plano z usando a transformação bilinear. | 29 |
| Figura 5 - Decomposição <i>wavelet</i> em cinco níveis. | 32 |
| Figura 6 - Exemplos de funções <i>wavelet</i> | 33 |
| Figura 7 - <i>Pipeline</i> metodológico para análise não supervisionada de sinais IM-BCI. | 45 |
| Figura 8 - Representação estrutural da matriz de características extraídas de sinais de EEG. | 58 |
| Figura 9 - ARI por sujeito para o <i>pipeline</i> de referência. | 63 |
| Figura 10 - Visualização dos agrupamentos obtidos por <i>k-means</i> na projeção dos dois primeiros PC para diferentes sujeitos. | 64 |
| Figura 11 - Visualização dos agrupamentos obtidos por <i>k-means</i> na projeção dos três primeiros PC para diferentes sujeitos. | 65 |
| Figura 12 - Comparação do ARI médio por etapa do pré-processamento | 67 |
| Figura 13 - Visualização bidimensional dos agrupamentos utilizando os pares canal-característica mais frequentemente selecionados..... | 71 |
| Figura 14 - Visualização tridimensional dos agrupamentos utilizando os pares canal-característica mais frequentemente selecionados..... | 72 |
| Figura 15 - Frequência normalizada de seleção dos canais do EEG nas combinações avaliadas. | 75 |
| Figura 16 - Frequência de seleção das características do EEG nas combinações avaliadas. | 76 |

LISTA DE TABELAS

| | |
|---|----|
| Tabela 1 - Polinômios de Butterworth normalizados | 28 |
| Tabela 2 - Notação para a tabela de contingência para comparar duas partições | 43 |
| Tabela 3 - Tabela de especificação dos eletrodos | 46 |
| Tabela 4 - Diferentes opções de pré-processamento..... | 57 |
| Tabela 5 – ARI médio para as combinações de filtro e janela temporal sem normalização | 68 |
| Tabela 6 – ARI médio para as combinações de filtro e janela temporal utilizando normalização <i>z-score</i> | 69 |
| Tabela 7 – ARI médio para as combinações de filtro e janela temporal utilizando normalização min-máx | 70 |
| Tabela 8 - Top 10 pares canal-característica por frequência de seleção. | 73 |

LISTA DE ABREVIATURAS E SIGLAS

| | |
|--------|---|
| ARI | <i>Adjusted Rand index</i> (índice de Rand ajustado) |
| BCI | <i>Brain-computer interface</i> (Interface cérebro-computador) |
| CFS | <i>Correlation-based Feature Selection</i> |
| DBSCAN | <i>Density-Based Spatial Clustering of Applications with Noise</i> |
| DWT | <i>Discrete wavelet transform</i> (transformada wavelet discreta) |
| EEG | Eletroencefalograma |
| EMG | Eletromiografia |
| ERD | <i>Event-related desynchronization</i> (dessincronização relacionada ao evento) |
| ERS | <i>Event-related synchronization</i> (sincronização relacionada ao evento) |
| IM | Imagética motora |
| PC | <i>Principal components</i> (componentes principais) |
| PCA | <i>Principal component analysis</i> (análise de componente principal) |
| TSFEL | <i>Time Series Feature Extraction Library</i> |

LISTA DE SÍMBOLOS

| | |
|-----------------|---|
| x_i | Amostra individual |
| \tilde{x}_i | Amostra normalizada (min-max) |
| \hat{x}_i | Amostra normalizada (<i>z-score</i>) |
| μ | Média |
| σ | Desvio padrão |
| X | Matriz de dados |
| $y[n]$ | Sinal de saída discreto |
| $x[k]$ | Sinal de entrada discreto |
| $h[n - k]$ | Resposta ao impulso invertida e deslocada |
| ε | Constante de atenuação máxima na banda passante |
| A_p | Atenuação na banda passante |
| n | Ordem do filtro |
| A_r | Atenuação na banda de rejeição |
| Ω | Frequência analógica |
| Ω_p | Frequência de corte da banda passante |
| Ω_s | Frequência de corte da banda de rejeição |
| $H(s)$ | Função de transferência no domínio de Laplace |
| $B_n(s)$ | Polinômio de Butterworth com ordem n |
| s | Frequência complexa normalizada |
| s_r | Frequência complexa no domínio físico |
| Ω_c | Frequência de corte analógica |
| T | Período de amostragem |
| z | Frequência complexa discreta |
| ω_c | Frequência de corte digital |
| $\Psi_{a,b}(t)$ | <i>Wavelet</i> -mãe |
| a | Variável de escala |
| b | Variável de translação |
| t | Tempo |
| k | Variável de translação discreta |
| j | Nível de decomposição |
| $m_{j,k}$ | Coefficientes <i>wavelet</i> |
| $x(t)$ | Sinal de entrada contínuo |

| | |
|-------------------|---|
| E_j | Energia <i>wavelet</i> do nível j |
| E_{tot} | Energia <i>wavelet</i> total |
| p_j | Energia <i>wavelet</i> relativa |
| S_{WT} | Entropia <i>wavelet</i> |
| A_x | Atividade de Hjorth |
| M_x | Mobilidade de Hjorth |
| C_x | Complexidade de Hjorth |
| $\dot{x}(t)$ | Primeira derivada do sinal de entrada contínuo |
| $\ddot{x}(t)$ | Segunda derivada do sinal de entrada contínuo |
| Cov | Matriz de covariância |
| X^T | Matriz de dados transposta |
| λ | Autovalor |
| v | Autovetor |
| P_i | Proporção da variância |
| Z | Escores dos componentes principais |
| J | Função objetivo |
| K | Número de <i>clusters</i> |
| C_j | Conjunto de dados atribuídos ao <i>cluster j</i> |
| c_j | Centroide atual do <i>cluster j</i> |
| D | Distância euclidiana |
| c_j^{novo} | Centroide novo do <i>cluster j</i> |
| u | Classes reais |
| w | <i>Clusters</i> encontrados |
| $q_{i,j}$ | Quantidade de objetos pertencentes à classe i e ao <i>cluster j</i> |
| $H(z)$ | Função de transferência no domínio discreto |
| f_i | Vetor de características |
| F | Matriz de características |
| ϕ | Vetor de valores de uma característica |
| ℓ | Vetor de rótulos de classe |
| ρ | Coefficiente de correlação de Pearson |
| Q_s | Mérito do subconjunto S |
| κ | Número de características no subconjunto |
| $\rho_{\phi\ell}$ | Correlação média entre característica e classe |
| $\rho_{\phi\phi}$ | Correlação média entre pares de características |

SUMÁRIO

| | | |
|----------------|--------------------------------------|-----------|
| 1 | INTRODUÇÃO | 15 |
| 1.1 | Objetivos | 16 |
| 1.1.1 | Objetivo geral | 16 |
| 1.1.2 | Objetivos específicos | 16 |
| 1.2 | Justificativa | 16 |
| 1.3 | Estrutura do trabalho | 17 |
| 2 | FUNDAMENTAÇÃO TEÓRICA | 18 |
| 2.1 | Imagética motora | 18 |
| 2.2 | Eletroencefalograma | 19 |
| 2.3 | Pré-processamento de sinais | 24 |
| 2.3.1 | Normalização | 24 |
| 2.3.2 | Filtragem de sinais | 25 |
| 2.4 | Extração de características | 30 |
| 2.4.1 | Transformada wavelet | 31 |
| 2.4.2 | Parâmetros de Hjorth | 35 |
| 2.5 | Redução de dimensionalidade | 36 |
| 2.5.1 | Análise de componentes principais | 37 |
| 2.6 | Aprendizado de máquina | 39 |
| 2.6.1 | Classificação por agrupamento | 40 |
| 2.6.2 | K-means | 40 |
| 2.6.3 | Índice de Rand ajustado | 42 |
| 3 | MATERIAIS E MÉTODOS | 44 |
| 3.1 | Descrição da base de dados | 45 |
| 3.1.1 | Participantes | 46 |
| 3.1.2 | Aquisição dos sinais de EEG | 46 |
| 3.1.3 | Procedimento experimental | 47 |
| 3.1.4 | Estrutura e formato de dados | 47 |
| 3.2 | <i>Pipeline</i> de referência | 48 |
| 3.2.1 | Pré-processamento | 48 |
| <u>3.2.1.1</u> | <u>Segmentação dos dados</u> | <u>48</u> |
| <u>3.2.1.2</u> | <u>Filtragem</u> | <u>49</u> |
| <u>3.2.1.3</u> | <u>Normalização</u> | <u>50</u> |
| <u>3.2.1.4</u> | <u>Seleção temporal</u> | <u>51</u> |

| | | |
|----------------|---|-----------|
| 3.2.2 | Extração de características | 51 |
| <u>3.2.2.1</u> | <u>Decomposição wavelet.....</u> | <u>52</u> |
| <u>3.2.2.2</u> | <u>Cálculo das métricas</u> | <u>52</u> |
| <u>3.2.2.3</u> | <u>Composição do vetor de características.....</u> | <u>53</u> |
| 3.2.3 | Redução de dimensionalidade | 54 |
| <u>3.2.3.1</u> | <u>Aplicação da análise de componentes principais</u> | <u>54</u> |
| <u>3.2.3.2</u> | <u>Critério de seleção dos componentes</u> | <u>55</u> |
| 3.3 | <i>Pipeline</i> de meta-aprendizagem | 55 |
| 3.3.1 | Pré-processamento | 56 |
| 3.3.2 | Extração de características | 57 |
| 3.3.3 | Seleção de canais e características | 58 |
| 3.4 | Agrupamento e avaliação de desempenho | 60 |
| 3.4.1 | Algoritmo k-means..... | 60 |
| 3.4.2 | Validação via índice de Rand ajustado..... | 61 |
| 4 | RESULTADOS E DISCUSSÕES | 62 |
| 4.1 | <i>Pipeline</i> de referência | 62 |
| 4.2 | <i>Pipeline</i> de meta-aprendizagem | 66 |
| 4.3 | Análise das características selecionadas | 72 |
| 5 | CONSIDERAÇÕES FINAIS | 77 |
| | REFERÊNCIAS | 79 |

1 INTRODUÇÃO

Dentro do cenário brasileiro, 18,6 milhões de pessoas possuem algum tipo de deficiência (Gomes, 2024). Essas condições impactam atividades diárias, como a locomoção e a comunicação. Por isso, o tema da acessibilidade para pessoas com deficiência é amplamente debatido. Neste contexto, o objetivo da acessibilidade é adaptar o ambiente para incluir pessoas com diferentes deficiências: contudo, não tem como intuito restaurar as funções perdidas, sejam elas presentes desde o nascimento ou fruto de fatalidades.

Ademais, o desafio enfrentado por pessoas com deficiência no âmbito familiar e no ambiente de trabalho é evidente. De acordo com Gomes (2024), o nível de ocupação — percentual de pessoas ocupadas na população em idade de trabalhar — foi de apenas 26,6% entre pessoas com deficiência. Ao restringir o recorte a aquelas que possuem nível superior, o índice de participação na força de trabalho sobe para 54,7%. Condições como paraplegia e tetraplegia são comuns no Brasil, e afetam sobretudo jovens adultos em sua fase mais produtiva da vida.

A fisioterapia exerce um papel central na reabilitação de pessoas com paraplegia e tetraplegia; no entanto, seus resultados são observados no ganho de resistência física e de força (Andrade, 2019). Dessa forma, a imagética motora (IM) é uma possibilidade para a reabilitação de funções motoras parcialmente ou totalmente comprometidas. A IM consiste na representação de um ato motor internamente, sem que haja qualquer manifestação externa (Souto, 2016). O ato de imaginar a execução de uma ação faz com que o cérebro emita sinais e, com base em uma metodologia de classificação, é possível elaborar estratégias de reabilitação multifacetadas para integrar pessoas com deficiência física à sociedade.

Para obter as atividades cerebrais, o método mais indicado é o eletroencefalograma (EEG), pois consiste em posicionar eletrodos no couro cabeludo do indivíduo e registrar os sinais sem interferir nas atividades cerebrais, o que elimina a necessidade de procedimentos invasivos ou de causar qualquer desconforto ou dor. Em relação a outros métodos, há três principais vantagens do EEG em estudos de IM. A primeira vantagem está associada ao método não ser invasivo e, portanto, permitir a realização de experimentos com sujeitos humanos normais. Outra vantagem do EEG é a resolução temporal alta, na ordem de milissegundos, que permite acompanhar a dinâmica temporal dos processos cerebrais. Por fim, o método é

relativamente barato, o que facilita a aplicação em clínicas e contextos de menor infraestrutura (Freeman; Quiroga, 2013).

As informações obtidas por meio da classificação dos sinais cerebrais podem ser aplicadas no desenvolvimento de próteses avançadas, com o objetivo de otimizar o sistema, melhorando o tempo de resposta e a acurácia. Ao compreender os padrões de ativação cerebral associados à imaginação de movimentos específicos, é possível projetar dispositivos capazes de traduzir essas intenções em ações físicas. Portanto, este trabalho conceber um sistema para a classificação de IM por meio de sinais de EEG.

1.1 Objetivos

1.1.1 *Objetivo geral*

Desenvolver um sistema inteligente não supervisionado para identificação de IM por meio de sinais de EEG.

1.1.2 *Objetivos específicos*

- Realizar uma revisão da literatura sobre classificação de IM e aplicação de algoritmos não supervisionados em sinais de EEG;
- Estudar as características do banco de dados experimental a ser utilizado;
- Implementar e avaliar sistematicamente técnicas de pré-processamento dos sinais de EEG, investigando o impacto de diferentes configurações de filtragem, normalização e janelamento temporal na discriminação das classes de IM;
- Extrair características nos domínios tempo-frequência, estatísticos e espectrais dos sinais de EEG e selecionar as características mais discriminativas entre as classes de IM por meio de um algoritmo, identificando os eletrodos e tipos de características com maior relevância populacional;
- Desenvolver e comparar dois sistemas não supervisionados de agrupamento de IM: um *pipeline* de referência e uma abordagem de meta-aprendizagem para identificar a configuração com melhor desempenho.

1.2 Justificativa

A relevância social deste trabalho justifica-se pelo propósito de apoiar parcela da população cuja qualidade de vida e autonomia são prejudicadas por

limitações motoras (Vaz, 2016). Diferentemente de alternativas como a fisioterapia convencional ou as medidas de acessibilidade, a proposta aqui é desenvolver soluções voltadas à reabilitação de funções motoras comprometidas. Tecnologias avançadas, como exoesqueletos e próteses robóticas de alto desempenho, embora promissoras, ainda apresentam custos elevados e exigem infraestrutura sofisticada, o que limita seu acesso em grande parte do território nacional. Nesse cenário, a aplicação de sistemas de classificação de IM por meio de EEG, aliada à portabilidade e à operação simplificada, amplia significativamente as possibilidades de uso em programas de reabilitação distribuídos, inclusive em áreas com acesso restrito a tecnologias avançadas.

A literatura científica internacional tem demonstrado resultados promissores na aplicação de técnicas de classificação de sinais de IM para controle de próteses e dispositivos assistivos. No entanto, observa-se uma lacuna significativa de estudos práticos conduzidos no contexto brasileiro, sobretudo no que diz respeito ao desenvolvimento de sistemas adaptados à nossa realidade socioeconômica (Brand; Santos; Midorikawa, 2018).

Diante disso, o desenvolvimento de um sistema de classificação representa não apenas uma contribuição científica, mas também uma oportunidade concreta de melhoria na qualidade de vida de pessoas com deficiência física, com potencial para fundamentar futuras aplicações clínicas e o desenvolvimento de soluções mais acessíveis e eficientes em reabilitação motora.

1.3 Estrutura do trabalho

O presente trabalho está organizado da seguinte maneira: no Capítulo 2, é apresentada a fundamentação teórica necessária para a definição dos objetos de estudo. A seguir, o Capítulo 3 detalha os procedimentos utilizados para a realização deste trabalho. O Capítulo 4 apresenta os resultados e as discussões sobre algoritmos implementados para cada tarefa de classificação. Por fim, o Capítulo 5 sumariza a pesquisa e apresenta as conclusões obtidas ao longo do desenvolvimento deste trabalho.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta os fundamentos teóricos que sustentam essa pesquisa. Discutem-se os principais conceitos sobre IM, seus correlatos neurofisiológicos, os padrões de ativação cerebral característicos e os métodos computacionais voltados à classificação de tarefas motoras imaginadas a partir de sinais cerebrais de EEG. Este referencial integra conhecimento das neurociências, do processamento de sinais e da ciência da computação, articulando-os na construção de sistemas de interface cérebro-computador (*Brain-Computer Interface*, BCI) capazes de interpretar a intenção motora dos usuários. São abordados também o panorama do EEG, os princípios de funcionamento da IM, as técnicas de extração de características dos sinais cerebrais e suas possibilidades de classificação por meio de métodos de aprendizagem não supervisionada, com ênfase em algoritmos de agrupamento e redução de dimensionalidade.

2.1 Imagética motora

O conceito de IM refere-se a um estado no qual o processo de mentalizar e reproduzir uma ação motora é representada internamente, isto é, o sujeito imagina a realização da ação, mas não há qualquer resposta motora associada (Souza et al., 2021). Há duas abordagens na IM: imagética visual e imagética cinestésica. A primeira envolve uma perspectiva em terceira pessoa, na qual o indivíduo se baseia em processamento visual e espacial ao assistir à execução da ação por outra pessoa (Decety, 1996). A segunda abordagem trata sob uma ótica em primeira pessoa, em que o sujeito busca sentir como se o movimento estivesse sendo executado, a fim de obter as sensações relacionadas às contrações musculares (Stecklow; Infantosi; Cagy, 2007).

A IM não necessita de estímulos externos, apenas das atividades cerebrais provenientes das intenções do indivíduo em executar a tarefa; dessa forma a prática mental tem o propósito de melhorar as condições motoras a partir do treinamento da IM, ou seja, a repetição sistemática de um movimento a partir de uma imagem mental (Nascimento, 2018). Nesse contexto, a IM é uma ferramenta terapêutica essencial na recuperação motora e performance funcional de indivíduos em reabilitação após lesão neurológica (Caldas et al., 2018).

Como o processo de imaginação depende principalmente dos mecanismos de processamento do sistema nervoso central, e não da capacidade física de executar a ação, pacientes com lesão cerebral unilateral tendem a demorar mais para visualizar mentalmente o movimento do membro afetado, em comparação ao membro não afetado. Por outro lado, indivíduos com lesão medular, paraplegia ou tetraplegia conseguem apresentar tempos de imaginação motora semelhantes aos observados em sujeitos saudáveis (Lameira et al., 2008; Lotze; Halsband, 2006). Em relação ao tempo de execução de determinado movimento, demonstrou-se que a duração da ação imaginada reproduz, de forma aproximada, a do mesmo movimento quando executado de fato. Essa correspondência temporal indica que a IM preserva propriedades dinâmicas fundamentais do ato motor. Por essa razão, também se observa que a IM segue a relação entre velocidade e precisão descrita pela Lei de Fitts, segundo a qual a duração do movimento está linearmente associada ao índice de dificuldade da tarefa (Decety, 1996).

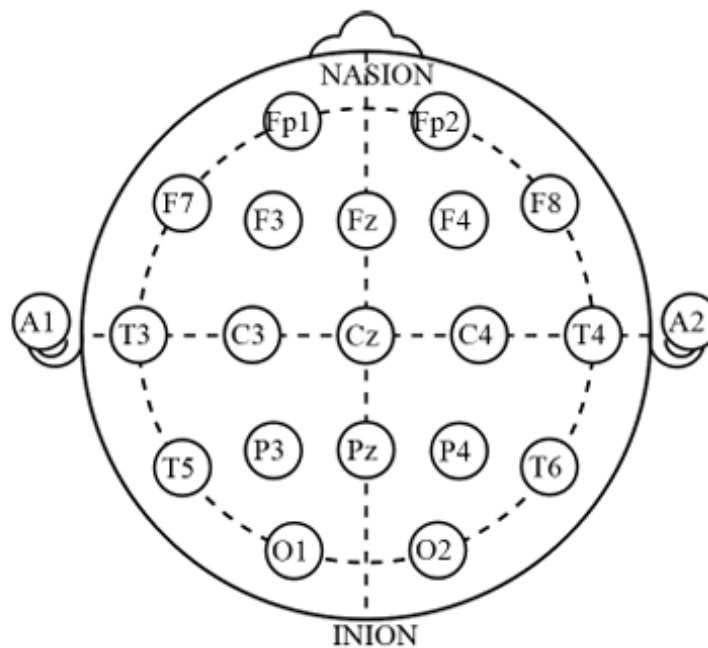
Considerando o potencial de aplicação da IM em neuropróteses e neuroreabilitação, destaca-se a BCI como componente central: trata-se de um canal artificial que estabelece comunicação entre o cérebro e o ambiente externo por meio de sistemas computacionais. O papel da BCI é captar e traduzir a atividade cerebral em sinais de controle capazes de operar computadores ou dispositivos assistivos. Essas respostas neurais podem ser obtidas por diferentes modalidades de aquisição, como ressonância magnética funcional, espectroscopia funcional no infravermelho e EEG (Oikonomou et al., 2017). Assim, a BCI pode prover um novo canal de comunicação para substituir funções motoras comprometidas ou possibilitar sistemas de resposta binária simples que auxiliem pacientes, promovendo ganhos de autonomia e de qualidade de vida. (Venkatachalam et al., 2020).

2.2 Eletroencefalograma

Atualmente, os estudos que envolvem o funcionamento do cérebro estão em desenvolvimento constante, de tal forma que as técnicas de registro vêm se aprimorando no mesmo ritmo. A aquisição de dados e diferentes desenhos de eletrodos permitem o registro simultâneo de inúmeros neurônios em maior escala, aliados ao avanço computacional possibilitam uma análise complexa dentro do que seria o campo de neurociência computacional (Freeman; Quiroga, 2013).

A obtenção dos sinais é realizada por meio de eletroencefalografia, que utiliza eletrodos de baixa impedância (alta condutividade) distribuídos pelo couro cabeludo em pontos específicos da cabeça como indicado na Figura 1. Esses pontos correspondem ao chamado sistema 10-20, nesse sistema 20 eletrodos são distribuídos por uma distância entre 10% e 20% da distância total referente à circunferência da cabeça (Lage, 2013). A atividade comum de um par de eletrodos, introduzidos nos lóbulos das orelhas (A1 e A2), serve como referência. As letras maiúsculas indicam os diferentes pontos dos eletrodos: C (central), F (frontal), O (occipital), P (parietal) e T (temporal). Números ímpares correspondem ao lado esquerdo e números pares ao lado direito, a letra z está associada à linha média central (Freeman; Quiroga, 2013).

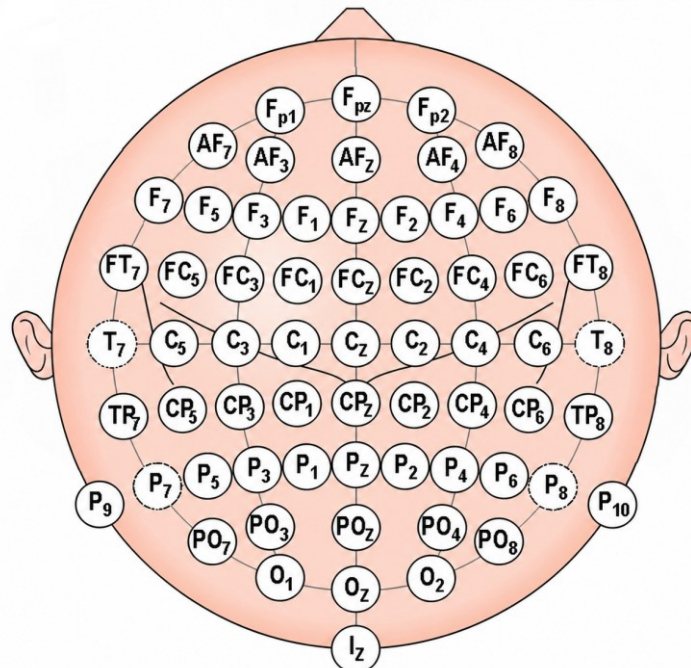
Figura 1 - Distribuição dos eletrodos no sistema 10-20.



Fonte: Lage (2013)

Uma extensão mais detalhada do sistema 10-20 é o sistema 10-10, no qual eletrodos adicionais são posicionados em regiões intermediárias entre os eletrodos convencionais, proporcionando maior resolução espacial e permitindo uma aquisição mais precisa das informações presentes nos sinais de EEG (Acharya et al., 2016). A disposição dos eletrodos no sistema 10-10 pode ser observada na Figura 2.

Figura 2 - Distribuição dos eletrodos no sistema 10-10 com 64 canais.



Fonte: Adaptado de Korats et al. (2013)

Os potenciais registrados no EEG são provenientes da circulação da condução iônica durante a excitação sináptica dos dendritos em uma enorme quantidade de neurônios próximos às camadas do córtex cerebral, portanto, o EEG consiste na aquisição das atividades elétricas nas quais os potenciais elétricos possuem amplitudes inferiores a 300 μV (Adur, 2008). Para indivíduos neurologicamente saudáveis, a atividade cortical demonstrada pelo EEG indica que a IM e a execução do movimento possuem padrões de potenciais relacionados a evento (Silva, 2018).

Um ponto importante a considerar é que os sinais antes de serem coletados pelo eletrodo ricocheteiam no cérebro; devido a isso, a resolução espacial é prejudicada, mas a resolução temporal é excelente. As ondas cerebrais apresentam frequências na faixa de 0,5 a 100 Hz; sistemas mais sofisticados facilmente podem registrar EEGs na faixa de 500 Hz. As bandas do EEG são definidas por faixas de frequência e associadas a estados como o sono, a atenção e patologias. As bandas são classificadas conforme a faixa de frequência, sendo elas vinculadas a uma letra grega (Adur, 2008). Segundo Freeman e Quian Quiroga (2013), existem seis padrões de ondas de EEG diferenciáveis:

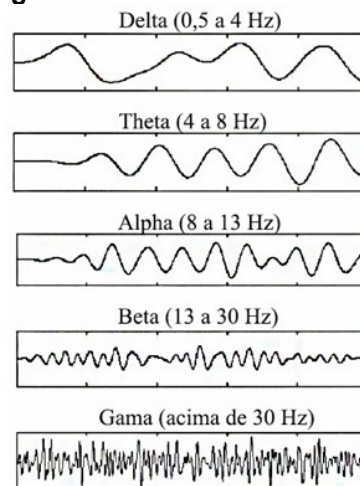
- Banda delta (0–4 Hz): características dos estágios profundos do sono. Conforme sua morfologia, localização e ritmo, as oscilações delta podem ser normais (como no sono de ondas lentas) ou patológicas (por exemplo, em tumores cerebrais).
- Banda theta (4–8 Hz): característica do sono profundo e desempenha papel importante na infância. Em adultos acordados, a alta atividade theta é anormal e pode indicar distúrbios cerebrais, como a epilepsia.
- Banda alfa (8–13 Hz): surge espontaneamente em adultos normais durante a vigília, em condições de relaxamento e inatividade mental.
- Banda mu: (8–13 Hz): possui frequência semelhante à banda alfa, porém com localização e função distintas. A banda mu é característica das regiões sensório-motoras centrais e é bloqueada especificamente durante movimentos, preparação motora ou observação de ações realizadas por outros, diferentemente da alfa que responde principalmente a estímulos visuais.
- Banda beta (13–30 Hz): são mais bem definidas em regiões centrais e frontais, com amplitude inferior à das ondas alfa. A sua amplitude aumenta durante esforços mentais abrangendo grande parte da superfície do couro cabeludo.
- Banda gama baixa (30-60 Hz): correlações da atividade gama com diferentes processos sensoriais e cognitivos, notadamente durante percepções visuais, auditivas, somáticas e olfativas, bem como com a atenção.
- Banda gama alta (definida de forma variável): também chamada de banda épsilon, foram observadas na escala *Eastern Cooperative Oncology Group* humana e animal em associação com potenciais de ação de rápida sucessão.

A análise da variação das potências absolutas em determinados ritmos cerebrais e em seus padrões de ondas, observados na Figura 3, está associada ao grau de sincronização das populações neurais. Estudos indicam que ocorrem alterações nos ritmos corticais durante o processo de estímulo de IM, sendo a diminuição ou o aumento na atividade cortical em bandas específicas, se comparadas a um período pré-evento, de sincronização relacionada ao evento (*Event-Related Synchronization*, ERS) e de dessincronização relacionada ao evento (*Event-Related Desynchronization*, ERD), respectivamente.

A ocorrência de ERD e ERS está diretamente relacionada aos mecanismos de excitação e inibição das populações neurais durante a execução de movimentos. Durante tarefas de IM, observa-se uma redução na amplitude dos ritmos alfa e beta do EEG em regiões do córtex sensório-motor contralateral ao movimento imaginado. De maneira geral o a ERD se manifesta antes e durante a execução do movimento, enquanto após o término da tarefa ocorre um aumento substancial da ERS (Silva, 2018).

Em tarefas envolvendo a IM da mão esquerda ou direita, os eletrodos contralaterais à tarefa apresentam redução de potência nas bandas alfa e beta durante a imaginação. Dessa forma, as diferenças topológicas entre as tarefas podem ser observadas principalmente nos eletrodos C3, C4 e Cz, posicionados sobre regiões do córtex sensório-motor (Yuan et al., 2010).

Figura 3 - Padrões de onda EEG



Fonte: Adaptado de Rashed-Al-Mahfuz et al. (2013)

A atividade relacionada ao movimento e à IM não se restringe ao córtex motor primário, mas envolve regiões corticais amplamente distribuídas, incluindo áreas frontoparietais, regiões frontais e temporais, áreas occipitais e o cerebelo pósterolateral (Yuan et al., 2010). Essa distribuição espacial ajuda a explicar por que, em análises de EEG aplicadas à IM, além dos canais centrais tradicionalmente associados ao córtex sensório-motor outros eletrodos também podem ser relevantes, dependendo da janela temporal analisada. Investigações baseadas em EEG demonstram, por exemplo, que o esforço de simulação mental e o planejamento do exercício induzem um aumento significativo nas potências de theta e alfa em áreas anteriores e de integração sensorial (Zhao et al., 2026). Isso se traduz no surgimento

de componentes críticos em canais frontais bem delimitados no tempo, refletindo o controle executivo frontal que opera em paralelo ou em antecedência à ativação das representações motoras puras localizadas nos canais centrais (Van der Lubbe et al., 2021).

Devido à sua baixa amplitude, os sinais de EEG são comumente afetados por fontes externas. Os chamados artefatos são todo potencial elétrico oriundo de fontes distintas do cérebro. Alguns tipos de artefatos resultam de atividades musculares, e outros provêm de componentes, como o 60 Hz da rede elétrica. Os artefatos comprometem os resultados da análise espectral e após o processamento, dificilmente podem ser removidos. Como alternativa para a correção desse problema podem ser aplicadas técnicas de pré-processamento como filtros ou técnicas de separação de fontes, preservando assim as faixas espectrais desejadas e excluindo as componentes indesejadas (Anghinah et al., 2006; Freeman; Quiroga, 2013; Korats et al., 2013).

2.3 Pré-processamento de sinais

Com o intuito de identificar padrões no EEG são aplicadas técnicas que extraem características para uma dada atividade específica. No caso de sinais bioelétricos como o EEG que são suscetíveis à interferência externa existe uma contaminação por ruídos, entre eles ruídos brancos, interferência da rede elétrica e outros artefatos. A análise dos sinais é prejudicada pela instabilidade e pela irregularidade. Para garantir estimativas precisas e melhor interpretação, o pré-processamento dos dados é essencial em algoritmos de extração de características, em que o principal objetivo é remover os ruídos e artefatos (Huang; Wang, 2021; Melo Junior et al., 2021).

2.3.1 Normalização

Um desafio recorrente no uso de bases de dados experimentais é a variabilidade das faixas de valores. Desse modo, tirar conclusões a partir de dados com diversas faixas de valores prejudica qualquer análise de um sistema dependente dessas informações. Uma abordagem eficaz para tratar essa questão consiste em transformar as variáveis dos dados em um espaço alternativo, assegurando que permaneçam necessariamente dentro de intervalos comparáveis. A normalização,

também denominada padronização de dados, constitui uma etapa fundamental do pré-processamento em sistemas computacionais. Esse procedimento é amplamente empregado em diversos campos, incluindo reconhecimento de padrões e aplicações desenvolvidas com *deep learning* (Akbulut, 2020).

A contribuição da normalização de dados em algoritmos de redes neurais que utilizam retropropagação é a simplificação dos cálculos numéricos, esse procedimento pode favorecer o aumento da eficiência do processo de aprendizado. Para a aprendizagem e classificação não supervisionada o impacto da normalização dos dados é considerado como uma transformação de dados na qual as distribuições se tornam mais normais e homogêneas, no entanto, devido à homogeneidade dos dados ocorre uma menor separabilidade entre as classes (Akbulut, 2020).

Existem diversos métodos para realizar a normalização em bases de dados, entre eles: min-máx, decimal e normalização pela norma. A normalização min-máx trata-se de uma transformação linear que permite preservar o formato do sinal, onde uma amostra x_i é convertida em \tilde{x}_i conforme a Equação 1:

$$\tilde{x}_i = \frac{x_i - \min(X)}{\max(X) - \min(X)}, \quad (1)$$

onde $\max(X)$ e $\min(X)$ representam o valor máximo e mínimo de um conjunto de dados, respectivamente. No entanto, a utilização desse método ocasiona a compressão dos dados quando existem valores discrepantes dentro do sinal, resultando em informações distorcidas (Zhu et al., 2025).

Portanto, para preservar ao máximo as características do sinal, o método *z-score* é amplamente utilizado para normalização. A técnica consiste em padronizar uma amostra (x_i) para média (μ) zero e desvio padrão (σ) unitário (Zhu et al., 2025). O efeito é equalizar as características com maior e menor amplitude, sendo definido pela Equação 2:

$$\hat{x}_i = \frac{x_i - \mu}{\sigma}. \quad (2)$$

2.3.2 Filtragem de sinais

Os filtros têm como objetivo principal selecionar componentes de frequência, rejeitando, assim, os indesejados de um sinal de entrada. O projeto de um filtro inicia-se a partir das especificações das propriedades desejadas para o sistema a ser desenvolvido, para então implementar o sistema, porém, isso só ocorre em sistemas

ideais. Em casos reais, é necessária uma aproximação às especificações utilizando um sistema discreto e causal. Espera-se que um filtro, ao processar um sinal de entrada, apresente amplitude unitária ou ganho constante na faixa de frequências a ser preservada no espectro, denominada banda passante. As faixas a serem eliminadas devem, por sua vez, receber ganho zero, de modo a anular sua contribuição (Oppenheim; Schafer, 2010).

A utilização de filtros com seletividade de frequência decorre da necessidade de separar os ruídos presentes no sinal analisado. Por exemplo, se o ruído em uma gravação de áudio estiver em uma banda de frequência mais alta do que a da voz, ele pode ser removido por filtragem seletiva em frequência. Embora a natureza das frequências a serem passadas por um filtro seletivo de frequência varie conforme a aplicação, vários tipos básicos de filtro são amplamente utilizados. As bandas passantes e de rejeição são definidas por limites de frequência estabelecidos pela frequência de corte. Os tipos de filtros seletivos de frequência mais comuns são o filtro passa-baixas, o passa-altas e o passa-faixa. O passa-baixas corresponde a um filtro que passa as frequências inferiores à frequência de corte da banda passante. Ao contrário do passa-baixas, o passa-altas incorpora à banda passante as frequências acima da frequência de corte. Por fim, o passa-faixa é um filtro caracterizado por atenuar as frequências entre os limites da banda passante; os valores dos limites são denominados de frequência de corte inferior e superior (Oppenheim; Willsky; Nawab, 1997).

Uma ferramenta fundamental na análise de sinais em tempo discreto é a convolução, pois, para sistemas lineares e invariantes no tempo, ela descreve completamente como um sinal de entrada é transformado pelo sistema. Para esse tipo de sistema qualquer sinal pode ser interpretado como a soma ponderada de impulsos unitários deslocados no tempo. Cada um desses impulsos gera uma resposta característica do sistema, e, devido às propriedades de linearidade e invariância temporal, a saída total resulta da superposição dessas respostas individuais. Cada amostra do sinal de entrada $x[k]$ atua como um fator de ponderação que multiplica uma versão deslocada da resposta ao impulso do sistema, e a soma dessas contribuições individuais compõe o sinal de saída $y[n]$. Dessa forma, a resposta ao impulso se torna uma representação completa do comportamento do sistema, possibilitando prever sua atuação sobre qualquer entrada. Além disso, esse princípio estabelece a base conceitual para o projeto e implementação de filtros

digitais, bem como para diversas operações envolvidas no processamento de sinais em tempo discreto (Oppenheim; Willsky; Nawab, 1997). Esse comportamento de um sistema linear invariante no tempo (LTI) é formalizado matematicamente por meio da soma de convolução discreta, conforme expresso na Equação 3:

$$y[n] = \sum_{k=-\infty}^{\infty} x[k]h[n-k], \quad (3)$$

em que $h[n-k]$ representa a resposta ao impulso do sistema invertida temporalmente e deslocada no tempo por k amostras, determinando como o sistema propaga a influência de cada pulso de excitação ao longo do tempo.

Os filtros com resposta ao impulso de duração infinita são sistemas cuja resposta a uma excitação se estende indefinidamente no tempo devido à presença de realimentação em sua estrutura. Esse tipo de filtro é obtido a partir de uma função de transferência expressa como a razão de dois polinômios, o que introduz polos e torna o comportamento dependente dos estados anteriores da saída. Uma característica importante dos filtros com resposta ao impulso de duração infinita, em relação aos filtros com resposta ao impulso de duração finita, é a capacidade de atender especificações de magnitude com ordens relativamente baixas, o que reduz o custo computacional e os torna adequados para aplicações em tempo real. Por outro lado, exigem atenção quanto à estabilidade e à resposta de fase, já que a presença de polos pode introduzir distorções temporais que devem ser consideradas no projeto (Diniz; Da Silva; Netto, 2010).

O filtro de Butterworth, descrito originalmente por Stephen Butterworth em 1930, é uma das topologias de filtro mais utilizadas no processamento de sinais. Ele é caracterizado por apresentar uma resposta de magnitude maximamente plana na banda de passagem, o que implica a ausência de ondulações tanto na faixa de passagem quanto na de rejeição. Devido a essa característica, a resposta em frequência é monotônica, decaindo suavemente para zero à medida que a frequência tende ao infinito (Schaumann; Xiao; Van Valkenburg, 2011).

No processo de aproximação e especificação de filtros, o comportamento da magnitude da resposta em frequência é determinado por parâmetros de projeto fundamentais. O primeiro deles é o fator de ondulação, calculado por meio da Equação 4,

$$\epsilon = \sqrt{10^{0,1A_p} - 1}, \quad (4)$$

em que ϵ serve para quantificar a atenuação ou variação máxima permitida no ganho dentro da banda de passagem em função da perda máxima especificada A_r (em dB). A partir dessa relação e das exigências estipuladas para a banda de rejeição, determina-se a ordem mínima do filtro por meio da Equação 5,

$$n \geq \frac{\log_{10} \left(\frac{10^{0,1A_r} - 1}{\epsilon^2} \right)}{2 \log_{10} \left(\frac{\Omega_s}{\Omega_p} \right)}, \quad (5)$$

em que n representa o número de polos, ou também a ordem do sistema, necessários na estrutura matemática do filtro, servindo diretamente para ditar a complexidade computacional do sistema e a inclinação da faixa de transição necessária para atingir a atenuação mínima A_r a partir das frequências limites da banda de passagem Ω_p e de rejeição Ω_s .

Uma vez determinada a ordem inteira mínima n , o fator de ondulação ϵ deixa de ser necessário explicitamente, pois sua função foi estabelecer o valor de n . A função de transferência de um filtro de Butterworth normalizado é então construída a partir da localização de seus polos no plano complexo de Laplace, distribuídos no semiplano esquerdo para garantir a estabilidade do sistema, sendo expressa em termos dos polinômios de Butterworth $B_n(s)$ conforme a Equação 6,

$$H(s) = \frac{1}{B_n(s)}, \quad (6)$$

$B_n(s)$ é o polinômio de Butterworth de ordem n , cujo índice corresponde diretamente ao valor calculado pela Equação 5. Os polinômios normalizados de primeira a quinta ordem são apresentados na Tabela 1.

| Tabela 1 - Polinômios de Butterworth normalizados | |
|--|---|
| Ordem (n) | Polinômio $B_n(s)$ |
| 1 | $s + 1$ |
| 2 | $s^2 + 1,4142s + 1$ |
| 3 | $(s + 1)(s^2 + s + 1)$ |
| 4 | $(s^2 + 0,7654s + 1)(s^2 + 1,8478s + 1)$ |
| 5 | $(s + 1)(s^2 + 0,6180s + 1)(s^2 + 1,6180s + 1)$ |

Fonte: Adaptado de Schaumann (2011).

Os polinômios de Butterworth tabelados na literatura consideram uma frequência de corte angular normalizada $\Omega_c = 1 \text{ rad/s}$. Para aplicar esses polinômios a um filtro real com uma frequência de corte específica Ω_c , utiliza-se a propriedade de escalonamento de frequência. A transformação do domínio normalizado (s) para o

domínio físico (s_r) em um filtro passa-baixa é feita pela substituição descrita na Equação 7 nos polinômios normalizados e pela reorganização dos termos para que o coeficiente de maior ordem de s seja unitário.

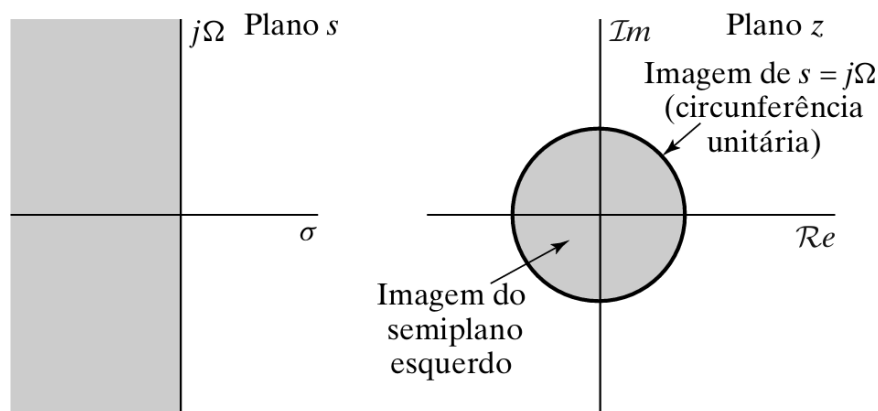
$$s = \frac{s_r}{\Omega_c}. \quad (7)$$

Para a implementação do filtro de Butterworth em sistemas digitais, utiliza-se frequentemente o método da transformação bilinear (Oppenheim; Schafer, 2010). Este método realiza um mapeamento do plano s para o plano z para um determinado tempo de amostragem (T), preservando a estabilidade e as características de magnitude do filtro analógico. A Equação 8 descreve a relação necessária para realizar o mapeamento.

$$s_r = \frac{2(z-1)}{T(z+1)}. \quad (8)$$

A Figura 4 ilustra o processo de mapeamento realizado por meio da transformação bilinear. O plano s , representado por um semiplano à esquerda que garante a estabilidade do sistema analógico, é mapeado para o interior da circunferência unitária no plano z . Dessa forma, o eixo imaginário $j\Omega$ do plano s é transformado na circunferência unitária do plano z , preservando as características de estabilidade do filtro digital resultante. Este mapeamento é fundamental para garantir que polos estáveis no domínio analógico permaneçam estáveis no domínio digital (Oppenheim; Schafer, 2010).

Figura 4 - Mapeamento do plano s no plano z usando a transformação bilinear.



Fonte: Oppenheim et al. (2010).

Apesar de preservar a estabilidade, a transformação bilinear introduz um mapeamento não linear entre a frequência angular analógica (Ω) e a frequência angular digital (ω). Essa não-linearidade resulta em uma distorção conhecida como deformação de frequência, na qual a escala de frequências analógicas é comprimida ao ser mapeada para o domínio digital. A relação entre as frequências é dada pela Equação 9.

$$\Omega_c = \frac{2}{T} \tan\left(\frac{\omega_c T}{2}\right). \quad (9)$$

Para compensar essa distorção e garantir que frequências críticas (como a frequência de corte de um filtro Butterworth) sejam mapeadas corretamente para o valor desejado no sistema digital, utiliza-se a técnica de pré-distorção. O procedimento consiste em calcular uma nova frequência analógica projetada, com base na frequência digital de corte desejada, antes de realizar o projeto do filtro analógico. Dessa forma, ao aplicar a transformação bilinear, a frequência de corte é ajustada para a posição correta no domínio digital.

2.4 Extração de características

A análise automatizada de sinais de EEG é uma tarefa complexa devido à natureza não-estacionária do sinal e à presença de artefatos que podem causar interferências durante a captura. Dentro do fluxo de classificação de um sistema de diagnóstico auxiliar, a etapa de extração de características desempenha um papel crucial após o pré-processamento (De Almeida; Lima; Peres, 2021).

A extração de características consiste em aplicar uma transformação às variáveis de entrada originais para projetar um vetor de alta dimensionalidade em um espaço de baixa dimensionalidade, resultando em um conjunto de novas variáveis no espaço transformado. O objetivo principal desta etapa é extrair as informações mais relevantes, discriminantes e não redundantes para a tarefa de classificação, facilitando a identificação de padrões intrínsecos ao sinal. É importante distinguir esta etapa da seleção de características: enquanto a seleção busca um subconjunto das variáveis originais, a extração estabelece atributos que podem não ter um significado físico isolado, mas que preservam a informação essencial com a mínima perda.

2.4.1 Transformada *wavelet*

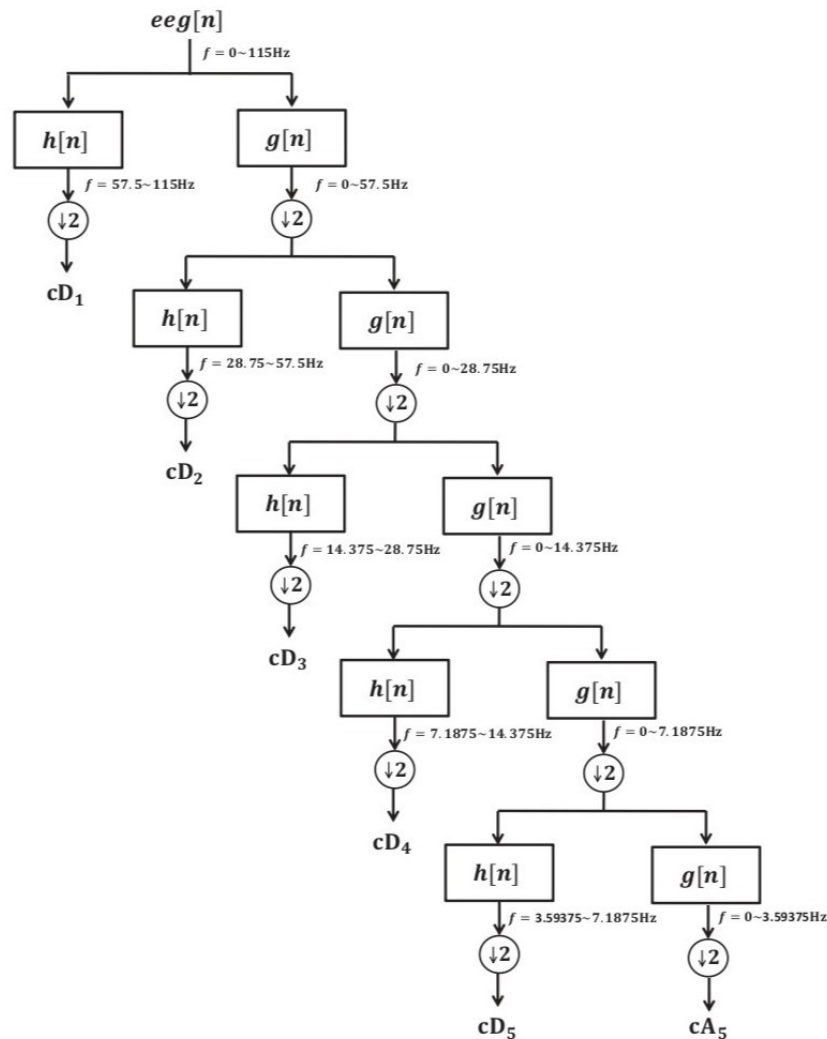
Neste contexto, as técnicas de domínio tempo-frequência emergem como cruciais para a análise de sinais não estacionários como o EEG. Especificamente, a transformada *wavelet*, com suas variantes como a transformada *wavelet* discreta (DWT), destaca-se por sua capacidade intrínseca de reduzir a dimensionalidade e extrair componentes relevantes do sinal simultaneamente, sendo uma ferramenta robusta tanto para a extração de características quanto para o pré-processamento visando a eficiência dos classificadores (De Almeida; Lima; Peres, 2021).

A transformada *wavelet* pode ser encarada como uma extensão da transformada clássica de Fourier, na qual, em vez de atuar apenas numa única escala (tempo ou frequência), opera em várias escalas. Essa capacidade multiescalar permite decompor um sinal em diferentes níveis, cada um representando um grau de detalhe do sinal analisado. A Figura 5 representa um exemplo de decomposição via transformada *wavelet* em um sinal de EEG, isto é, para uma faixa de frequência base é realizada uma divisão dela em subfaixas menores com o objetivo de isolar uma faixa de frequência específica. A decomposição em escalas é especialmente vantajosa quando a transformada é aplicada sobre uma base ortogonal. Assim, as componentes de alta frequência são interpretadas como ruído e podem ser eliminadas. É precisamente essa capacidade de fracionar o sinal em várias escalas e descartar partes indesejadas que torna a transformada *wavelet* uma técnica atraente para o processamento de sinais. (Hazarika et al., 1997).

A *wavelet* é uma função matemática oscilatória, em que o sinal é decomposto em versões "escaloadas" e "transladadas" de uma função base única, a *wavelet*-mãe $\psi(t)$. Os valores de escala (a) controlam a dilatação ou contração da *wavelet*, ou seja, valores menores comprimem a onda, enquanto valores maiores a esticam. A variável (b) está associada à translação, isto é, o parâmetro que define o deslocamento da *wavelet* ao longo do tempo, permitindo saber quando o evento ocorreu (Gandhi; Panigrahi; Anand, 2011). A Equação 10 expressa o princípio básico dessa transformação.

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \Psi\left(\frac{t-b}{a}\right), \quad a > 0, b \in \mathbb{R} \quad (10)$$

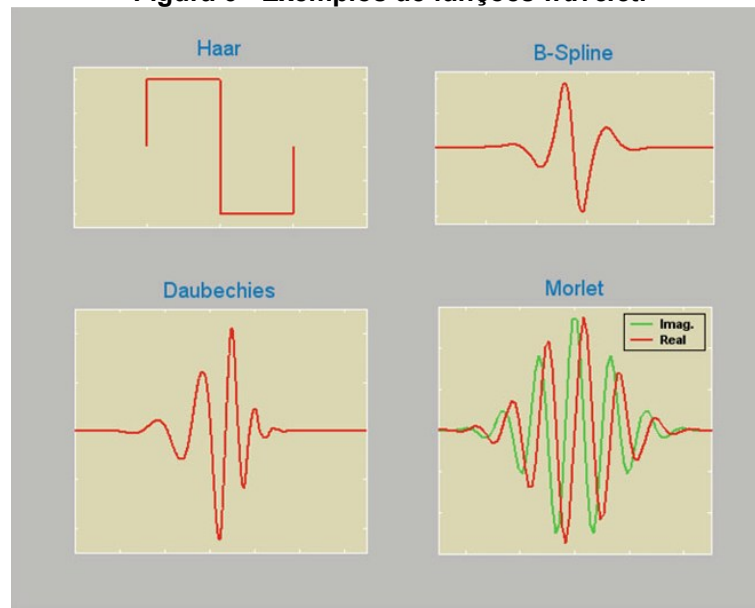
Figura 5 - Decomposição *wavelet* em cinco níveis.



Fonte: Vivas *et al.* (2013).

Vale ressaltar que existem muitas funções diferentes que podem ser utilizadas como *wavelet*, e a possibilidade de escolher uma *wavelet*-mãe a partir de um dicionário de funções possíveis é uma das principais forças dessa técnica. Cada uma dessas funções apresenta características distintas, cuja importância depende, tipicamente, da aplicação em mente. A Figura 6 ilustra quatro funções amplamente utilizadas: a Haar, caracterizada por sua simplicidade e forma de onda quadrada; a B-Spline, que apresenta maior suavidade; a Daubechies, conhecida por seu suporte compacto e assimetria; e a Morlet, que se assemelha a uma senoide modulada. Essa flexibilidade permite selecionar a forma de onda que melhor se correlaciona com as estruturas do sinal analisado.

Figura 6 - Exemplos de funções *wavelet*.



Fonte: Freeman e Quiroga (2013, p. 53).

A DWT é obtida ao restringir os valores que os parâmetros a e b podem assumir, transformando-os de variáveis contínuas em discretas. O método mais comum é a amostragem diádica, que baseia a discretização em potências de dois. Nesse método assume-se que os parâmetros de escala e de translação são $a = 2^j$ e $b = k2^j$, respectivamente. Onde $j \in \mathbb{Z}$ indica o nível de decomposição e $k \in \mathbb{Z}$ a posição temporal (Chen et al., 2017). Essa estrutura é formalmente definida pela Equação 11.

$$\Psi_{j,k}(t) = 2^{-\frac{j}{2}}\Psi(2^{-j}t - k). \quad (11)$$

Nesta formulação, cada incremento na variável j resulta na duplicação da escala a . Isso implica dobrar a largura da janela de análise no domínio do tempo e reduzir pela metade a largura no domínio da frequência, resultando em uma análise por bandas de oitava. Ao fixar a base de discretização, é possível selecionar uma *wavelet* mãe tal que as funções $\Psi_{j,k}(t)$ formem uma base ortonormal, permitindo a reconstrução exata do sinal original por meio da soma ponderada de seus coeficientes $m_{j,k}$ conforme a Equação 12 (Diniz; Da Silva; Netto, 2010).

$$x(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} m_{j,k} \Psi_{j,k}(t), \quad (12)$$

sendo os coeficientes $m_{j,k}$ calculados pela Equação 13.

$$m_{j,k} = \int_{-\infty}^{\infty} \Psi_{j,k}(t) x(t) dt. \quad (13)$$

A análise direta dos coeficientes *wavelet* pode ser complexa devido à grande quantidade de dados. Para sintetizar essa informação e caracterizar o sinal, a literatura utiliza medidas derivadas da teoria da informação e processamento de energia. As duas principais métricas são a energia e a entropia *wavelet*. Dado que a família de funções *wavelet* $\Psi_{j,k}(t)$ forma uma base ortonormal para o espaço de funções de quadrado integrável $L^2(\mathbb{R})$, o conceito de energia pode ser diretamente vinculado ao Teorema de Parseval (Rosso et al., 2001).

A energia do sinal em um nível de resolução específico j é calculada pelo somatório do quadrado dos coeficientes de detalhe correspondentes conforme a Equação 14.

$$E_j = \sum_k |m_{j,k}|^2. \quad (14)$$

Consequentemente, a energia total do sinal (E_{tot}) é a soma das energias de todos os níveis de decomposição, como pode ser observado na Equação 15.

$$E_{tot} = \sum_j \sum_k |m_{j,k}|^2. \quad (15)$$

A partir desses valores, define-se a energia *wavelet* relativa p_j (ver Equação 16), que normaliza a energia de cada nível em relação à energia total. Essa distribuição comporta-se como uma distribuição de probabilidade da energia por meio das escalas (densidade tempo-escala). Isso fornece uma ferramenta adequada para detectar e caracterizar fenômenos específicos nos planos de tempo e frequência.

$$p_j = \frac{E_j}{E_{tot}}. \quad (16)$$

A entropia *wavelet* combina a decomposição *wavelet* com a entropia de Shannon para fornecer uma medida de variabilidade associada às diferentes bandas de frequência. Ela funciona como um critério para analisar e comparar a distribuição de probabilidade descrita anteriormente (Rosso et al., 2001; Sharma; Pachori; Acharya, 2015). Utilizando a distribuição de energia relativa, a entropia *wavelet* total (S_{WT}) é definida pela Equação 17.

$$S_{WT} = - \sum_j p_j \log_2 p_j. \quad (17)$$

A entropia quantifica o grau de organização de um sinal, funcionando como um indicador da sua dinâmica subjacente, ou seja, em processos ordenados, a energia concentra-se predominantemente em um único nível de resolução, resultando em uma entropia *wavelet* mínima, próxima de zero. Por outro lado, em processos aleatórios ou desordenados, a energia distribui-se de forma uniforme entre todas as bandas de frequência, fazendo com que a entropia *wavelet* atinja seus valores máximos.

2.4.2 Parâmetros de Hjorth

Um desafio enfrentado na análise de um sinal de EEG é a necessidade de métodos quantitativos para sua avaliação, de forma a estabelecer parâmetros adequados. Segundo Hjorth (1970), para desenvolver tais parâmetros, é necessário reconhecer a diferença entre duas abordagens principais: a caracterização geral de um padrão de amplitude e tempo por meio de métodos estatísticos; e a detecção de padrões individuais previamente definidos. Ao focar na primeira abordagem, busca-se extrair descritores que resumam matematicamente o comportamento global do sinal sem a necessidade de procurar por anomalias específicas.

Nesse contexto, os parâmetros de Hjorth surgem como descritores estatísticos do domínio do tempo capazes de resumir, de forma compacta, as propriedades globais do EEG sem recorrer à identificação de padrões específicos. Esses parâmetros são divididos em três medidas complementares: atividade, mobilidade e complexidade. A atividade, que corresponde à variância da amplitude do sinal e pode ser interpretada como potência média; mobilidade, que expressa a relação entre a variabilidade da derivada de primeira ordem e a variabilidade do próprio sinal, refletindo a rapidez média de suas variações; e complexidade, que compara a mobilidade da derivada do sinal com a mobilidade do sinal original, quantificando o grau de irregularidade em relação a uma forma senoidal ideal, para a qual esse valor mínimo é igual a 1 (Hjorth, 1970). Dessa forma, esses parâmetros permitem caracterizar o EEG em termos de amplitude, escala temporal e complexidade estrutural, conforme apresentado nas Equações 18, 19 e 20:

$$A_x = \sigma_x^2, \quad (18)$$

$$M_x = \frac{\sigma_{\dot{x}}}{\sigma_x} = \sqrt{\frac{\text{var}(\dot{x}(t))}{\text{var}(x(t))}}, \quad (19)$$

$$C_x = \frac{\frac{\sigma_{\dot{x}}}{\sigma_x}}{\frac{\sigma_{\ddot{x}}}{\sigma_{\dot{x}}}} \quad (20)$$

em que $x(t)$ representa o sinal EEG, $\dot{x}(t)$ sua primeira derivada temporal e $\ddot{x}(t)$ sua segunda derivada temporal.

Devido à natureza não-estacionária do EEG e à forte variabilidade interindividual durante tarefas de IM, a extração de características é um desafio crítico. Nesse sentido, os Parâmetros de Hjorth têm sido amplamente adotados por condensarem eficientemente as propriedades temporais e espectrais do sinal. A literatura comprova que a parametrização destas métricas otimiza o desempenho de sistemas BCI, aumentando a precisão de classificação em 4,4% ao adaptar o modelo ao perfil de ativação de cada utilizador (Oh; Lee; Kim, 2014). Adicionalmente, em aplicações clínicas críticas, a Atividade e a Complexidade de Hjorth destacam-se pela sua notável estabilidade algorítmica; apresentando os menores índices de dispersão e desvio padrão entre métodos de extração, estes parâmetros mitigam o ruído e asseguram consistência na detecção de intenção motora, mesmo em pacientes com graves distúrbios de consciência (Höller et al., 2013).

2.5 Redução de dimensionalidade

Frequentemente, o grande número de características envolvidas no aprendizado de máquina torna-se um problema. Além de atrasar significativamente o treinamento, esse excesso dificulta a busca por uma boa solução, fenômeno conhecido como “maldição da dimensionalidade”. Apesar desse obstáculo, em muitos casos é possível reduzir o número de variáveis para simplificar a resolução do problema. Essa redução de dimensionalidade acelera o treinamento e atua como um filtro, descartando ruídos e detalhes desnecessários.

É importante ressaltar, no entanto, que esse processo causa certa perda de informação, o que pode resultar em leve queda no desempenho do sistema. Por outro lado, um benefício crucial de reduzir as dimensões é a facilitação da visualização dos dados, permitindo a detecção de padrões e agrupamentos (*clusters*) de forma puramente visual (Géron, 2019). Para realizar essa compressão estruturada lidando com esse compromisso entre informação e viabilidade, diversas abordagens matemáticas podem ser aplicadas, abrindo espaço para técnicas como a Análise de Componente Principal (*Principal Component Analysis*, PCA).

2.5.1 Análise de componentes principais

Conjuntos de dados grandes são cada vez mais comuns e frequentemente difíceis de interpretar. Nesse contexto, a PCA surge como uma técnica estatística fundamental para redução de dimensionalidade, atuando de forma a aumentar a interpretabilidade, mas ao mesmo tempo minimizando a perda de informações (Jolliffe; Cadima, 2016).

A combinação de transformada *wavelet* com PCA tem-se mostrado particularmente eficaz no processamento de sinais EEG. Como os sinais EEG são não estacionários, o método convencional de análise de frequência não é altamente bem-sucedido na classificação diagnóstica, e a transformada *wavelet* é particularmente efetiva para representar vários aspectos de sinais não estacionários (Subasi; Erçelebi, 2005).

Para sistemas não supervisionados, esta integração é especialmente vantajosa. Foi demonstrado que a PCA é uma técnica útil para agrupamento, e em neurociência, a PCA é usada para reduzir a dimensionalidade do espaço de formas de onda de potenciais de ação, seguido por análise de agrupamento (Jolliffe, 2002). Ao reduzir o número de dimensões, a PCA simplifica os dados, preservando sua estrutura, tornando algoritmos de agrupamento, como *k-means* ou *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN), mais eficientes.

A PCA baseia-se em conceitos de álgebra linear e de estatística multivariada. Conforme descrito detalhadamente por Jolliffe (2002), o processo matemático pode ser dividido nas seguintes etapas: a determinação da matriz de covariância, os autovalores e autovetores, seleção dos Componentes Principais (*Principal Components*, PC) e cálculo dos escores.

Uma vez que o conjunto de dados X (de dimensão $n \times p$) foi devidamente normalizado o primeiro passo é obter a relação linear entre as variáveis. Neste cenário específico de dados padronizados, a matriz de covariância (Cov). coincide com a matriz de correlação. Sendo assim, ela pode ser obtida por meio de operações de álgebra matricial simples indicadas na Equação 21, onde n representa o número de observações.

$$Cov = \frac{1}{n-1} X^T X. \quad (21)$$

O resultado é uma matriz quadrada e simétrica de dimensão $p \times p$ (onde p é o número de variáveis), onde os elementos da diagonal principal são iguais a 1 e os demais elementos representam o coeficiente de correlação linear entre as variáveis.

Uma vez obtida a matriz de covariância, o passo seguinte consiste em realizar a sua decomposição espectral com o objetivo de encontrar os autovalores (λ) e os autovetores (v) que satisfazem a Equação 22.

$$(Cov - \lambda)v = 0. \quad (22)$$

É essencial que os autovalores sejam ordenados de forma decrescente ($\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$), de modo que o primeiro autovalor corresponda à direção de maior variância nos dados. Para cada autovalor λ_i , calcula-se o autovetor correspondente v_i , sendo que estes vetores determinam os pesos que cada variável original terá na composição do componente principal.

Com os autovalores calculados, deve-se decidir quantos PC (k) serão mantidos na análise. Como o objetivo da PCA é a redução de dimensionalidade, descartam-se os componentes que explicam pouca variância. Assim, a proporção da variância total explicada pelo i -ésimo componente principal é dada pela Equação 23. Geralmente, selecionam-se os primeiros componentes cuja soma acumulada da variância atinja um limiar pré-estabelecido, descartando-se os demais.

$$P_i = \frac{\lambda_i}{\sum_{j=1}^p \lambda_j}. \quad (23)$$

A etapa final do procedimento consiste em projetar os dados originais no novo espaço vetorial definido pelos componentes selecionados. Os novos valores (escores) para cada observação no i -ésimo componente principal (Z_i) são obtidos pela combinação linear das variáveis padronizadas (X) ponderadas pelos elementos do autovetor (v_i) como representado na Equação 24. Desta forma, o conjunto de dados original de dimensão p é transformado em um novo conjunto de dados de dimensão k (onde $k < p$), mantendo as informações mais relevantes e eliminando a redundância entre as variáveis.

$$Z_i = v_{i1}X_1 + v_{i2}X_2 + \dots + v_{ip}X_p. \quad (24)$$

2.6 Aprendizado de máquina

O aprendizado de máquina (*Machine Learning*) é descrito, fundamentalmente, como a capacidade de um agente melhorar seu desempenho em tarefas por meio de observações do mundo. Diferente da programação tradicional, onde regras são explicitamente codificadas, no aprendizado de máquina, segundo Russell (2021, p. 651), "o computador observa dados, constrói um modelo baseado nesses dados e usa esse modelo tanto como uma hipótese sobre o mundo quanto como um software capaz de resolver problemas".

A preferência pelos métodos de aprendizagem em vez do uso da programação direta de soluções é consequência da imprevisibilidade e da complexidade das soluções. A imprevisibilidade resulta de situações em que os projetistas de sistemas não conseguem antecipar todas as possíveis situações futuras. Por exemplo, um robô de navegação ou um algoritmo de previsão do mercado financeiro precisa se adaptar a novas configurações ou mudanças econômicas que não existiam no momento de sua criação. Em termos de complexidade de soluções, os humanos possuem a habilidade de realizar tarefas de forma subconsciente, mas não sabem explicar ou programar explicitamente como essa tarefa é realizada (Russell et al., 2021).

De acordo com Russell (2021), o aprendizado de máquina é classificado em três tipos principais, baseados no tipo de feedback disponível para o agente:

- **Aprendizado supervisionado:** O agente observa pares de entrada e saída e aprende uma função que mapeia a entrada para a saída. O objetivo é prever o rótulo correto para novas entradas. Isso pode ser dividido em classificação e regressão.
- **Aprendizado não supervisionado:** O agente aprende padrões a partir da entrada, sem receber feedback explícito ou rótulos. Uma tarefa comum é o agrupamento, em que o sistema identifica grupos de dados similares, como categorias de imagens, sem saber a priori o que elas representam.
- **Aprendizado por reforço:** O agente aprende por meio de uma série de reforços. Por exemplo, em um jogo, o agente pode receber uma recompensa apenas ao vencer. Ele deve então determinar quais ações anteriores foram responsáveis pelo sucesso e ajustar seu comportamento para maximizar as recompensas futuras.

2.6.1 *Classificação por agrupamento*

O agrupamento é a técnica predominante no paradigma do aprendizado não supervisionado. Nesse contexto, o agrupamento é definido como o processo de identificar estruturas latentes nos dados, atribuindo observações a grupos específicos de modo que pontos considerados "similares" pertençam ao mesmo grupo, enquanto pontos "diferentes" sejam segregados em grupos distintos (Sukhorukova; Kelly, 2024).

A relação dessa técnica com a análise de EEG é fundamental devido à complexidade e à natureza, muitas vezes não rotulada, desses sinais biológicos. Em aplicações clínicas, como a detecção de epilepsia, nem sempre é possível para especialistas identificarem visualmente e rotularem todos os estados cerebrais, especialmente os momentos sutis que antecedem uma convulsão. Pesquisas recentes demonstram que métodos não supervisionados são essenciais para descobrir "estados ocultos" nos dados sem anotações clínicas prévias. Ao aplicar algoritmos de agrupamento, é possível varrer grandes volumes de dados de EEG para isolar padrões de ondas cerebrais que caracterizam o estado pré-convulsivo, uma tarefa difícil de realizar via métodos supervisionados tradicionais que dependem de dados previamente classificados (Jiman; Attar, 2025).

Além da detecção de patologias, o agrupamento demonstra eficácia na distinção entre diferentes tarefas motoras e imaginárias. Análises de desempenho indicam que algoritmos como o *k-means* e o agrupamento hierárquico apresentam resultados superiores na detecção de atividade motora real das mãos, com taxas de sucesso superiores a 80%. Já no contexto de tarefas motoras imaginárias das mãos, o algoritmo hierárquico se destaca com 34% de taxa de acerto, enquanto o agrupamento espectral (*spectral clustering*) revela-se mais eficiente para a detecção de tarefas motoras envolvendo os pés (Asanza; Pelaez; Loayza, 2017).

2.6.2 *K-means*

O *k-means* é amplamente reconhecido na literatura como um dos algoritmos de agrupamento mais populares e eficazes dentro da categoria de aprendizado não supervisionado (Ahmed; Seraj; Islam, 2020). Sua função primária é resolver problemas de agrupamento (*clustering*) por meio do particionamento de um conjunto de dados em K grupos distintos e não sobrepostos. A premissa fundamental do algoritmo é agrupar observações de tal forma que os dados dentro de um mesmo

cluster sejam o mais semelhantes possível entre si (alta similaridade intra-grupo), ao passo que os dados de *clusters* diferentes sejam distintos (baixa similaridade inter-grupo) (Sukhorukova; Kelly, 2024).

O algoritmo opera de forma iterativa para encontrar uma partição que minimize a função-objetivo, comumente definida como a soma dos erros quadráticos. Matematicamente, isso significa minimizar a distância euclidiana entre cada ponto de dados e o centro geométrico do seu respectivo *cluster*, denominado centroide (Sukhorukova; Kelly, 2024). Embora seja valorizado por sua simplicidade de implementação e eficiência computacional linear o *k-means* possui limitações que devem ser consideradas, como a necessidade de especificar o número de *clusters* a priori e a sensibilidade à escolha dos centroides iniciais, o que pode levar o algoritmo a convergir para ótimos locais em vez do ótimo global (Ahmed; Seraj; Islam, 2020).

O algoritmo *k-means* opera como um método de otimização iterativa. Do ponto de vista do cálculo, seu objetivo é particionar um conjunto de dados $X = \{x_1, x_2, \dots, x_n\}$ em K *clusters* distintos, minimizando a variância intra-*cluster*. A função matemática que rege todo o processo é a função objetivo. A Equação 25 busca encontrar os centroides que minimizam J .

$$J = \sum_{j=1}^K \sum_{i \in C_j} \|x_i - c_j\|^2. \quad (25)$$

O processo de cálculo começa com o algoritmo selecionando aleatoriamente K pontos do espaço de dados para servirem como os centroides iniciais (c_1, \dots, c_k) . A escolha aleatória é o método padrão, mas pode levar a mínimos locais. Variações como o *k-means++* utilizam um cálculo de probabilidade proporcional à distância para escolher centros iniciais mais espaçados, o que melhora a convergência.

Após a inicialização, o algoritmo calcula a proximidade geométrica de cada ponto de dados em relação a todos os centroides existentes. Para medir essa proximidade entre pontos e centroides, utiliza-se a Equação 26, que representa a distância euclidiana.

$$D(x_i, c_k) = \sqrt{\sum_{l=1}^d (x_{i,l} - c_{k,l})^2}. \quad (26)$$

A etapa de atribuição de pontos aos *clusters* no algoritmo *k-means* é formalmente definida pela regra segundo a qual cada amostra é associada ao

centroide mais próximo. Matematicamente, o conjunto C_j de elementos pertencentes ao *cluster* j é descrito como $C_j = \{x_i: \|x_i - c_j\| \leq \|x_i - c_k\|, \forall k \neq j\}$. Essa expressão indica que um dado x_i será incluído em C_j sempre que a distância entre ele e o centroide c_j for menor ou igual à distância entre essa mesma amostra e qualquer outro centroide c_k . Dessa forma, a partição dos dados ocorre de modo que cada ponto é alocado à região do espaço onde o centroide correspondente exerce a maior proximidade, o que equivale a minimizar a distância euclidiana entre ponto e centroide. Assim, a equação não só formaliza o critério de associação no *k-means*, mas também expressa a essência do algoritmo para agrupar elementos pela semelhança medida via proximidade geométrica (Ahmed; Seraj; Islam, 2020; Sukhorukova; Kelly, 2024).

Após a atribuição, os centroides antigos deixam de representar o centro geométrico real dos novos grupos formados. É necessário recalcular a posição dos centroides. Sendo assim, a partir da Equação 27 o novo centroide c_j^{novo} é obtido calculando a média aritmética de todos os pontos atribuídos ao *cluster* C_j , movendo-o para o centro geométrico do grupo (Ahmed; Seraj; Islam, 2020; Izbicki; Dos Santos, 2020).

$$c_j^{novo} = \frac{1}{|C_j|} \sum_{x_j \in C_j} x_j. \quad (27)$$

A convergência do *k-means* é caracterizada pelo processo iterativo de minimizar a função da soma das distâncias euclidianas ao quadrado entre cada ponto de dados e o centroide do seu *cluster* correspondente. O algoritmo garante a convergência a um mínimo local da função de erro quadrático. Como a função objetivo decresce (ou permanece constante) a cada iteração e o número de configurações de partição possíveis é finito, o algoritmo inevitavelmente atinge um estado estacionário (Jain; Murty; Flynn, 1999).

2.6.3 Índice de Rand ajustado

O índice Rand ajustado (ARI) é uma métrica externa de validação de agrupamentos que mede a similaridade entre duas partições de um conjunto de dados. Em outras palavras, o ARI compara os pares de pontos que caem nos mesmos *clusters* em duas partições e ajusta esse valor pela similaridade esperada ao acaso. O ARI é definido de forma que valores próximos de +1 indicam concordância quase perfeita entre as partições, valores próximos de 0 correspondem à concordância

esperada por acaso, e valores negativos indicam concordância menor do que o acaso (De Assis, 2018).

A construção da Tabela 2 é o passo crítico para o cálculo do ARI: cada entrada q_{ij} indica o número de amostras que pertencem simultaneamente à classe verdadeira u_i e ao *cluster* w_j . A análise do ARI usa combinações de pares $\binom{q}{2}$ para computar a concordância observada entre as duas partições e corrige esse valor pelo esperado sob um modelo nulo de partições aleatórias, o que confere ao índice a propriedade desejável de ter expectativa zero quando não há estrutura além do acaso. Assim, o ARI é particularmente adequado para comparar resultados de algoritmos de agrupamento em relação a rótulos verdadeiros, pois permite comparações justas mesmo quando os tamanhos dos *clusters* variam. Estudos demonstram o uso prático do ARI em estudos de agrupamento combinados com redução de dimensionalidade por PCA, mostrando como a variação no número de componentes pode afetar sensivelmente a separabilidade medida pelo ARI (Yeung; Ruzzo, 2001).

Tabela 2 - Notação para a tabela de contingência para comparar duas partições

| Classe/Cluster | w_1 | w_2 | ... | w_C | Soma |
|----------------|----------|----------|-----|----------|--------------|
| u_1 | q_{11} | q_{12} | ... | q_{1C} | $q_{1.}$ |
| u_2 | q_{21} | q_{22} | ... | q_{2C} | $q_{2.}$ |
| ... | ... | ... | ... | ... | ... |
| u_R | q_{R1} | q_{R2} | ... | q_{RC} | $q_{R.}$ |
| Soma | $q_{.1}$ | $q_{.2}$ | ... | $q_{.C}$ | $q_{..} = q$ |

Fonte: Adaptado de Yeung e Ruzzo (2001).

Antes de calcular o índice, define-se também o número total de possíveis pares de elementos no conjunto, dado pela Equação 28:

$$\binom{q}{2} = \frac{q(q-1)}{2}. \quad (28)$$

Esse valor representa o número total de pares que podem ser formados a partir dos n objetos do conjunto. Assim, o ARI é calculado pela concordância observada entre as partições e subtraindo o valor esperado dessa concordância caso as partições fossem aleatórias. O resultado é normalizado para garantir que o índice varie entre -1 e $+1$, conforme mostrado na Equação 29:

$$ARI = \frac{\sum_{i,j} \binom{q_{ij}}{2} - [\sum_i \binom{q_{i.}}{2} \sum_i \binom{q_{.j}}{2}]}{\frac{1}{2} [\sum_i \binom{q_{i.}}{2} + \sum_i \binom{q_{.j}}{2}] - [\sum_i \binom{q_{i.}}{2} \sum_i \binom{q_{.j}}{2}]} \quad (29)$$

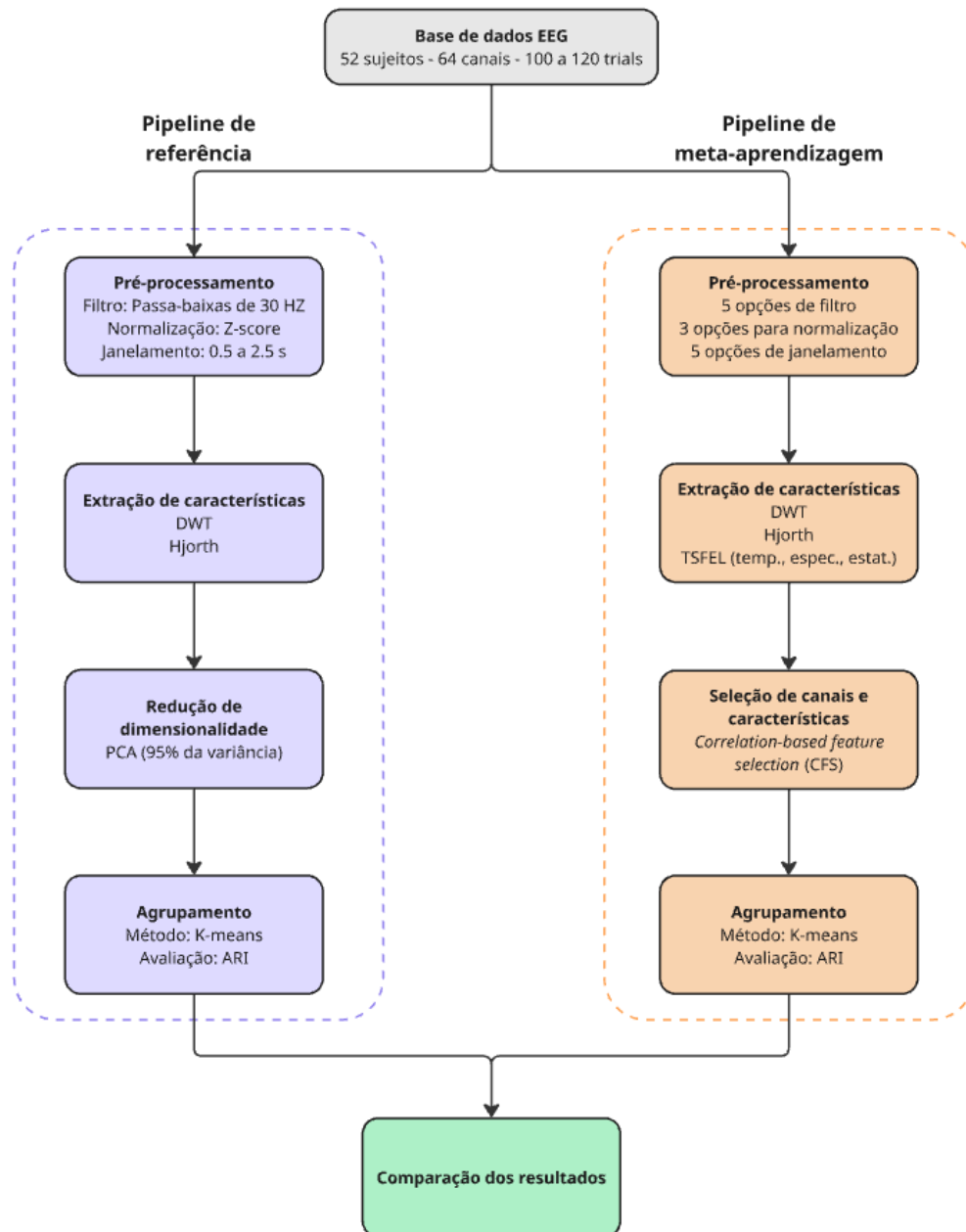
3 MATERIAIS E MÉTODOS

Este capítulo descreve a metodologia, visando explorar padrões intrínsecos nos dados que podem não corresponder diretamente às classes experimentais definidas (mão esquerda e mão direita), com o objetivo de compreender a organização natural desses sinais cerebrais. O *pipeline* metodológico, ilustrado na Figura 7, é estruturado em dois ramos paralelos e independentes: o de referência e o de meta-aprendizagem. Ambos compartilham a mesma base de dados e convergem para uma etapa final de comparação dos resultados.

A escolha por uma abordagem não supervisionada se justifica pela natureza complexa e multidimensional dos sinais de EEG, onde características ocultas e variações interindividuais podem não ser completamente capturadas por métodos tradicionais de classificação supervisionada, ou seja, sinais cerebrais em tarefas de IM são influenciados por diversos fatores cognitivos e fisiológicos que podem transcender a simples dicotomia experimental. Ao permitir que os próprios dados revelem sua estrutura organizacional, este trabalho busca identificar subgrupos naturais de padrões cerebrais, investigar a correspondência entre esses agrupamentos e as classes experimentais, e potencialmente elucidar fatores relacionados à variabilidade de desempenho em sistemas de BCI.

O *pipeline* de referência estabelece uma linha de base com configurações fixas e consolidadas na literatura, ao passo que o *pipeline* de meta-aprendizagem estende essa abordagem ao avaliar sistematicamente o impacto de diferentes decisões de pré-processamento sobre a qualidade do agrupamento. As seções seguintes detalham cada componente metodológico, incluindo a base de dados utilizada, os parâmetros específicos de cada técnica de processamento de sinais, os algoritmos de extração de características e de agrupamento empregados, bem como as métricas de avaliação adotadas para quantificar a qualidade, a interpretabilidade e a correspondência dos resultados obtidos com as classes experimentais.

Figura 7 - Pipeline metodológico para análise não supervisionada de sinais IM-BCI.



Fonte: Autoria própria (2025).

3.1 Descrição da base de dados

A base de dados utilizada neste trabalho consiste em um conjunto público de sinais de EEG para uma BCI baseada em IM, desenvolvido e disponibilizado no repositório GigaScience Database. Este conjunto de dados foi selecionado pela qualidade e completude dos dados, e por incluir metadados relevantes para investigação de fatores relacionados à variabilidade de desempenho entre sujeitos (Cho et al., 2017).

3.1.1 Participantes

O conjunto compreende dados de 52 sujeitos saudáveis (19 mulheres e 33 homens), com idade média de $24,8 \pm 3,86$ anos. Do total de participantes, 50 eram destros, e 2 eram ambidestros. Todos os sujeitos assinaram o termo de consentimento informado e o experimento foi aprovado pelo Comitê de Ética em Pesquisa do Instituto de Ciência e Tecnologia de Gwangju (*Gwangju Institute of Science and Technology*), Coreia do Sul.

3.1.2 Aquisição dos sinais de EEG

Os sinais de EEG foram registrados utilizando o sistema BioSemi ActiveTwo, com 64 eletrodos ativos de Ag/AgCl, posicionados de acordo com o sistema internacional 10-10, conforme detalhado na Tabela 3. A taxa de amostragem utilizada foi de 512 Hz. Adicionalmente, foram registrados sinais de eletromiografia (EMG) síncronos com o EEG, por meio de quatro eletrodos posicionados nos músculos flexor digitorum profundus e extensor digitorum de ambos os braços, permitindo a detecção de movimentos reais não intencionais durante as tarefas de IM.

Tabela 3 - Tabela de especificação dos eletrodos

| Nº do canal | Eletrodo | Nº do canal | Eletrodo | Nº do canal | Eletrodo | Nº do canal | Eletrodo |
|-------------|----------|-------------|----------|-------------|----------|-------------|----------|
| 1 | Fp1 | 17 | CP5 | 33 | Fpz | 49 | C2 |
| 2 | AF7 | 18 | CP3 | 34 | Fp2 | 50 | C4 |
| 3 | AF3 | 19 | CP1 | 35 | AF8 | 51 | C6 |
| 4 | F1 | 20 | P1 | 36 | AF4 | 52 | T8 |
| 5 | F3 | 21 | P3 | 37 | AFz | 53 | TP8 |
| 6 | F5 | 22 | P5 | 38 | Fz | 54 | CP6 |
| 7 | F7 | 23 | P7 | 39 | F2 | 55 | CP4 |
| 8 | FT7 | 24 | P9 | 40 | F4 | 56 | CP2 |
| 9 | FC5 | 25 | PO7 | 41 | F6 | 57 | P2 |
| 10 | FC3 | 26 | PO3 | 42 | F8 | 58 | P4 |
| 11 | FC1 | 27 | O1 | 43 | FT8 | 59 | P6 |
| 12 | C1 | 28 | Iz | 44 | FC6 | 60 | P8 |
| 13 | C3 | 29 | Oz | 45 | FC4 | 61 | P10 |
| 14 | C5 | 30 | POz | 46 | FC2 | 62 | O2 |
| 15 | T7 | 31 | Pz | 47 | FCz | 63 | PO4 |
| 16 | TP7 | 32 | CPz | 48 | Cz | 64 | PO8 |

Fonte: Autoria própria (2025).

Para cada sujeito, foram capturadas as coordenadas tridimensionais de localização de cada eletrodo utilizando um digitalizador Polhemus Fastrak, possibilitando análises espaciais precisas e a correção de variações individuais no posicionamento dos eletrodos.

3.1.3 *Procedimento experimental*

O procedimento experimental seguiu o protocolo padrão para tarefas de IM bilateral. Cada ensaio consistiu em três etapas: período de preparação, período de tarefa e período de descanso. Durante o período de preparação, uma tela preta com uma cruz de fixação era exibida por 2 segundos, indicando ao participante que um novo ensaio estava prestes a começar. Em seguida, ocorre um período de tarefa em que uma instrução visual aleatória ("mão esquerda" ou "mão direita") é apresentada por 3 segundos, durante o qual o participante deve imaginar o movimento da mão correspondente. Para finalizar é realizado um período de descanso onde é apresentada uma tela em branco por um intervalo aleatório entre 4,1 e 4,8 segundos, permitindo ao sujeito relaxar antes do próximo ensaio.

Os sujeitos foram instruídos a imaginar movimentos cinestésicos em vez de movimentos visuais. Especificamente, a tarefa consistia em imaginar o toque sequencial de cada dedo (indicador, médio, anelar e mínimo) no polegar no período de 3 segundos, conforme praticado previamente com movimentos reais.

3.1.4 *Estrutura e formato de dados*

Cada sujeito realizou entre 5 e 6 sessões, com cada uma contendo 20 ensaios por classe, totalizando entre 100 e 120 ensaios por classe e por sujeito. Após cada sessão, os sujeitos recebiam *feedback* sobre a acurácia de classificação obtida, com o objetivo de manter a motivação. Intervalos de até 4 minutos foram oferecidos entre as sessões, conforme solicitado pelos participantes.

Os dados estão organizados em arquivos MATLAB (.mat) individuais para cada sujeito, contendo estruturas com os sinais de EEG (canais 1-64) e EMG (canais 65-68), marcadores temporais de eventos, taxa de amostragem, coordenadas 3D dos eletrodos e metadados associados. Os sinais de EEG são fornecidos em formato contínuo, em que todos os ensaios de uma mesma classe estão concatenados em uma única matriz de dimensão $[68 \text{ canais} \times N_{amostras}]$. Para a classe de imagética de mão esquerda (*imagery_left*) e mão direita (*imagery_right*), cada matriz possui dimensão $[68 \times 358400]$ para sujeitos com 100 ensaios ou $[68 \times 430080]$ para sujeitos com 120 ensaios, correspondendo a 100 ou 120 segmentos de 7 segundos cada, respectivamente.

A segmentação dos sinais contínuos em ensaios individuais é realizada por meio do vetor de eventos de imagética (*imagery_event*), que possui a mesma dimensão temporal dos sinais e contém valores binários, onde "1" indica o instante de *onset* de cada ensaio (momento da apresentação da instrução visual) e "0" representa os demais instantes. Cada ensaio foi registrado em uma janela temporal de -2000 a +5000 ms em relação ao *onset*, totalizando 7 segundos por ensaio.

3.2 Pipeline de referência

O *pipeline* de referência tem como objetivo estabelecer uma linha de base para a análise não supervisionada dos sinais de EEG de IM, adotando configurações de pré-processamento e extração de características consolidadas na literatura de BCI. A metodologia segue um fluxo sequencial composto por quatro etapas: pré-processamento dos sinais brutos, extração de características no domínio tempo-frequência, redução de dimensionalidade e agrupamento por *k-means*.

O pré-processamento é realizado com parâmetros fixos de modo a isolar as bandas de frequência sensório-motoras e padronizar os sinais entre sujeitos e sessões. A extração de características combina a DWT com os parâmetros de Hjorth, resultando em um vetor compacto de 448 dimensões por ensaio. A PCA é então aplicada para reduzir a redundância espacial e mitigar a maldição da dimensionalidade, mantendo os componentes responsáveis por 95% da variância total. Por fim, o algoritmo *k-means* particiona os ensaios em dois agrupamentos, cujos rótulos são comparados às classes experimentais pelo ARI.

3.2.1 Pré-processamento

A etapa de pré-processamento tem como objetivo preparar os sinais brutos de EEG para a extração de características, removendo artefatos, ruídos e componentes de frequência irrelevantes, ao mesmo tempo em que padroniza a amplitude dos sinais. As etapas de pré-processamento implementadas consistem em filtragem seletiva em frequência e normalização estatística.

3.2.1.1 Segmentação dos dados

Os dados brutos de EEG são disponibilizados em formato contínuo, em que todos os ensaios de uma mesma classe são concatenados sequencialmente em uma

única matriz bidimensional. Para possibilitar a análise individualizada de cada ensaio, realizou-se a segmentação destes sinais utilizando os marcadores temporais fornecidos no vetor *imagery_event*.

Considerando que cada ensaio possui duração total de 7 segundos e que a taxa de amostragem é de 512 Hz, cada ensaio corresponde a um bloco de 3584 amostras. O procedimento de separação consistiu no particionamento da matriz contínua em N blocos de tamanho igual. Dessa forma, a matriz original de dimensão $[68 \times 358400]$ para 100 ensaios foi reorganizada em uma estrutura de dimensão $[64 \times 3584 \times 100]$, ou $[64 \times 3584 \times 120]$ para 120 ensaios, delimitando os dados aos 64 canais referentes ao EEG (descartando-se os 4 canais de EMG nas posições 65-68) e isolando cada repetição da tarefa para o processamento subsequente. Este procedimento foi aplicado separadamente às matrizes *imagery_left* e *imagery_right*, resultando em duas estruturas tridimensionais correspondentes às duas classes de IM.

3.2.1.2 Filtragem

Após a segmentação dos ensaios, cada segmento do sinal bruto foi submetido a um processo de filtragem digital para remover artefatos e isolar as bandas de frequência de interesse. Foi projetado um filtro digital do tipo Butterworth passa-baixas de 5ª ordem ($n = 5$), com frequência de corte digital (f_c) estabelecida em 30 Hz. A escolha desta frequência de corte fundamenta-se na necessidade de preservar os ritmos sensório-motores, especificamente as bandas mu e beta, que contêm as principais características neurofisiológicas associadas à IM, ao mesmo tempo em que se atenuam componentes de alta frequência provenientes de ruído instrumental e artefatos musculares. Para corrigir a distorção não linear da transformação bilinear, utilizou-se a pré-distorção da frequência de corte digital indicada na Equação 9, obtendo a frequência angular analógica equivalente de $\Omega_c = 190,6584 \text{ rad/s}$.

Partindo da função de transferência normalizada do filtro Butterworth de quinta ordem (Equação 6 e Tabela 1), aplicou-se o escalonamento de frequência por meio da substituição descrita na Equação 7, resultando na função de transferência analógica. Por fim, foi realizado o mapeamento para o domínio digital por meio da Equação 8 e expandidos algebricamente os termos da função de transferência digital representada pela Equação 30:

$$H(z) = \frac{0,0001228(z + 1)^5}{z^5 - 3,8111z^4 + 5,919z^3 - 4,668z^2 + 1,865z - 0,3013} \quad (30)$$

A função de transferência digital resultante foi implementada em MATLAB por meio da função *butter()*¹ do *Signal Processing Toolbox*, que realiza automaticamente as etapas de pré-distorção, desnormalização e transformação bilinear. Para garantir resposta de fase zero e evitar distorções temporais, aplicou-se filtragem bidirecional por meio da função *filtfilt()*², que processa o sinal nas direções direta e reversa, resultando em uma resposta efetiva de ordem 10 com fase linear. A escolha da filtragem bidirecional é apropriada para a análise de sinais de EEG, sendo amplamente utilizada por preservar as relações temporais dos padrões neurais, sem introduzir atrasos de fase.

3.2.1.3 Normalização

Dada a natureza não estacionária dos sinais de EEG e a significativa variabilidade de amplitude observada entre diferentes sujeitos e sessões, a presença de valores brutos de tensão poderia enviesar os algoritmos de reconhecimento de padrões. Para mitigar estas discrepâncias e garantir que todas as características possuam a mesma escala de importância, foi implementada a execução de um procedimento de normalização estatística.

Cada segmento filtrado de 7 segundos foi submetido, individualmente, à padronização por *z-score*. O cálculo fundamenta-se na subtração da média do segmento, seguida da divisão pelo seu desvio padrão, conforme a relação matemática descrita na Equação 2, apresentada no referencial teórico. A operacionalização deste procedimento foi realizada por meio da função *zscore()*³, integrante do *Statistics and Machine Learning Toolbox*. A escolha por esta rotina computacional deve-se à sua otimização para operações matriciais e precisão numérica. A função foi executada considerando a dimensão temporal da matriz de dados, o que assegura o cálculo independente das estatísticas para a série temporal de cada um dos 64 canais. Tal abordagem garante o respeito à dinâmica individual de cada eletrodo, preservando a topografia espacial do sinal de EEG, sem que ocorra a mistura de amplitudes entre canais distintos.

¹ Disponível em: <https://www.mathworks.com/help/signal/ref/butter.html>. Acesso em: 3 dez, 2025

² Disponível em: <https://www.mathworks.com/help/signal/ref/filtfilt.html>. Acesso em: 3 dez, 2025

³ Disponível em: <https://www.mathworks.com/help/stats/zscore.html>. Acesso em: 3 dez, 2025

A decisão metodológica de aplicar a normalização à janela estendida de 7 segundos, previamente à seleção do intervalo de interesse, tem por objetivo conferir maior robustez estatística. Ao serem considerados, no cálculo, tanto o período pré-estímulo quanto o período de atividade, a dinâmica relativa de potência do sinal é preservada.

3.2.1.4 Seleção temporal

Uma vez padronizados, embora os segmentos pré-processados possuam 7 segundos de duração, utilizar todo o período no classificador introduziria ruído não relacionado à tarefa cognitiva. Dessa forma, restringiu-se a análise a uma janela de interesse entre +0,5 s e +2,5 s após o início do estímulo visual. A escolha dessa janela de 2 segundos (1024 amostras) deve-se ao descarte dos primeiros 500 ms, contaminados pela influência dos potenciais evocados visuais gerados durante o processamento do estímulo. Além disso, esse intervalo abrange o tempo de reação e o planejamento motor, nos quais a modulação dos ritmos mu e beta (ERD) ainda não atingiu seu pico de estabilidade.

Ao final desta etapa de pré-processamento, obtém-se, para cada ensaio, uma matriz de dados de dimensão $[64 \times 1024]$, filtrada, normalizada e temporalmente focada. Estes dados encontram-se devidamente condicionados para a etapa subsequente de extração de características, na qual serão submetidos à decomposição em sub-bandas de frequência.

3.2.2 *Extração de características*

A etapa de extração de características tem por objetivo mapear os sinais pré-processados para um espaço de menor dimensionalidade que maximize a separabilidade entre as classes. Considerando que a IM induz alterações espectrais transientes (ERD/ERS) e que os sinais de EEG são inerentemente não estacionários. Sendo assim, foi utilizada a DWT para transformar os sinais do domínio do tempo para o domínio tempo-frequência, sobre as matrizes de dados previamente pré-processadas a fim de obter características relevantes para a discriminação de classe. Complementarmente a análise por DWT, foram extraídos os parâmetros de Hjorth.

3.2.2.1 Decomposição wavelet

A decomposição multirresolução foi operacionalizada por meio da função *wavedec()*⁴, integrante do *Wavelet Toolbox*. Esta rotina implementa o algoritmo de Mallat, realizando a filtragem e a subamostragem sucessivas para decompor o sinal em coeficientes de aproximação e de detalhe.

Para cada um dos 64 canais, a função foi executada definindo as 1024 amostras da janela temporal selecionada como entrada. Considerando a frequência de amostragem de 512 Hz, foi definido o uso da família Daubechies 4, devido à adequação morfológica aos potenciais de EEG, com um nível 5 de decomposição. Esta profundidade é necessária para que as sub-bandas de frequência resultantes isolem adequadamente os ritmos cerebrais de interesse nos coeficientes de detalhe dos níveis intermédios (D3, D4, D5) e na aproximação final (A5).

A execução da função resulta em dois vetores de saída: o vetor de coeficientes decompostos (*C*), que contém a concatenação de todos os coeficientes de aproximação e detalhe [*cA5, cD5, cD4, cD3, cD2, cD1*], e o vetor de contabilidade (*L*), que armazena o comprimento de cada segmento de coeficientes, permitindo a sua reconstrução ou extração individual.

3.2.2.2 Cálculo das métricas

A extração dos descritores estatísticos foi realizada a partir da estrutura de decomposição fornecida pela função *wavedec()*. Para assegurar a precisão numérica e a reprodutibilidade dos cálculos, a operacionalização foi realizada por meio de rotinas específicas da *Wavelet Toolbox* do MATLAB, com foco em duas propriedades fundamentais do sinal: a energia relativa e a entropia.

A energia relativa foi calculada a partir da função *wenergy()*⁵, que processa a estrutura de decomposição completa e retorna um vetor contendo a percentagem de energia associada à aproximação final e a cada nível de detalhe, conforme demonstrado na Equação 16. Esta abordagem normaliza a contribuição energética de cada sub-banda em relação à energia total do sinal, proporcionando uma medida invariante em relação à amplitude absoluta.

⁴ Disponível em: <https://www.mathworks.com/help/wavelet/ref/wavedec.html>. Acesso em: 3 dez, 2025

⁵ Disponível em: <https://www.mathworks.com/help/wavelet/ref/wenergy.html>. Acesso em: 3 dez, 2025

Simultaneamente, a entropia é calculada utilizando a função *wentropy()*⁶, aplicada individualmente aos coeficientes recuperados de cada nível. Esta medida permite identificar padrões não lineares e o grau de desordem da atividade neural, oferecendo uma característica complementar à energia, que captura aspectos da complexidade e regularidade dos ritmos cerebrais associados à IM.

Complementarmente, foram extraídos os três parâmetros de Hjorth diretamente a partir do sinal temporal janelado de cada canal, conforme as Equações 18, 19 e 20. As derivadas temporais foram aproximadas numericamente pela diferença finita de primeira ordem entre amostras consecutivas, de acordo com a implementação adotada para sinais digitais discretos.

3.2.2.3 Composição do vetor de características

Embora a decomposição *wavelet* de nível 5 gere coeficientes para todo o espectro de frequência, a construção do vetor de características foi orientada pelo conhecimento neurofisiológico da tarefa e pelas restrições de filtragem aplicadas anteriormente. Considerando que o filtro de pré-processamento limitou o sinal a 30 Hz e que a ERD na IM ocorre predominantemente nos ritmos mu e beta. Foram selecionados exclusivamente os coeficientes de detalhe dos níveis D4 (16-32 Hz) e D5 (8-16 Hz), que correspondem às bandas de interesse neurofisiológico.

As bandas de detalhe de alta frequência (D1, D2, D3) foram excluídas por se encontrarem na banda de rejeição do filtro passa-baixas, contendo apenas ruído residual e componentes espectrais já atenuados. A aproximação final (A5) e o nível de detalhe mais baixo não foram utilizados para evitar a introdução de artefatos de baixa frequência como EMG e movimentos oculares.

Desta forma, foram extraídas 7 características por canal: energia relativa de D4 (E_{D4}), entropia de D4 ($S_{WT,D4}$), energia relativa de D5 (E_{D5}), entropia de D5 ($S_{WT,D5}$), atividade (A_x), mobilidade (M_x) e complexidade (C_x). Para os 64 canais de EEG, o vetor de características de cada ensaio foi construído pela concatenação sequencial dessas métricas, conforme a Equação 31:

$$f_i = [E_{D4}^{(1)} \quad S_{WT,D4}^{(1)} \quad E_{D5}^{(1)} \quad S_{WT,D5}^{(1)} \quad A_x^{(1)} \quad M_x^{(1)} \quad C_x^{(1)} \quad \dots \quad M_x^{(64)} \quad C_x^{(64)}], \quad (31)$$

⁶ Disponível em: <https://www.mathworks.com/help/wavelet/ref/wentropy.html>. Acesso em: 3 dez, 2025

onde o sobrescrito (i) indica o canal correspondente, resultando em um vetor \mathbb{R}^M com $M = 64 \times 7 = 448$.

Considerando que cada sujeito realizou 100 ensaios por classe, a matriz final de características para entrada na PCA foi construída conforme apresentado na Equação 32, com dimensões $N \times M$, onde $N = 200$ ou 240 (em indivíduos com 120 ensaios):

$$F = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_{200} \end{bmatrix}. \quad (32)$$

Esta estruturação combina descritores tempo-frequência com descritores temporais globais, proporcionando uma representação complementar dos padrões de IM: enquanto as métricas DWT capturam a distribuição de energia e complexidade nas bandas mu e beta, os parâmetros de Hjorth caracterizam a amplitude, a rapidez de variação e a irregularidade do sinal no domínio do tempo.

3.2.3 Redução de dimensionalidade

Apesar de a seleção criteriosa das bandas de frequência ter reduzido o espaço de características, a matriz de entrada ainda apresenta dimensionalidade elevada em relação ao número de ensaios disponíveis por sujeito. Além disso, devido à natureza da condução de volume no couro cabeludo, existe uma forte redundância espacial, em que canais vizinhos tendem a apresentar comportamentos oscilatórios altamente correlacionados. Manter um número excessivo de variáveis, muitas das quais redundantes, pode prejudicar o desempenho dos algoritmos de agrupamento e aumentar o custo computacional.

3.2.3.1 Aplicação da análise de componentes principais

A redução de dimensionalidade foi operacionalizada por meio da função `pca()`⁷, integrante do *Statistics and Machine Learning Toolbox*. O algoritmo recebeu como argumento de entrada a matriz unificada F de dimensão $[200 \times 448]$, onde as linhas correspondem aos ensaios e as colunas às características tempo-frequência por canal.

⁷ Disponível em: <https://www.mathworks.com/help/stats/pca.html>. Acesso em: 3 dez, 2025

A execução da função realiza a decomposição em valores singulares da matriz de covariância dos dados centrados e retorna, entre outros parâmetros, os coeficientes dos PC (autovetores), os escores (representação dos dados no novo sistema de coordenadas) e a variância explicada por cada componente.

3.2.3.2 Critério de seleção dos componentes

A determinação da dimensionalidade final do subespaço (k) não foi fixada a priori, mas sim guiada pela representatividade estatística dos componentes. Para isso foi utilizado o vetor de saída *explained* fornecido pela função, que contém a porcentagem da variância total dos dados explicada por cada componente principal, ordenada de forma decrescente.

O critério de corte baseou-se na variância explicada acumulada, ou seja, realizou-se o somatório progressivo dos valores do vetor *explained*, selecionando-se os primeiros k componentes necessários para atingir o limiar de 95% da variância total. Este procedimento assegura a manutenção da estrutura global da informação, descartando apenas as dimensões associadas ao ruído residual e às redundâncias lineares de baixa significância.

A partir da definição do valor de k , a matriz de dados reduzida (Y) foi constituída selecionando-se apenas as primeiras k colunas da matriz de scores retornada pela função. O resultado é uma matriz de dimensão $[200 \times k]$, onde cada linha representa a projeção otimizada de um ensaio no espaço dos PC. Esta estrutura constitui a entrada final para o algoritmo de agrupamento.

3.3 **Pipeline de meta-aprendizagem**

Na aprendizagem de máquina, o termo meta-aprendizagem refere-se ao uso de resultados de múltiplos experimentos de aprendizagem para orientar a seleção ou configuração de modelos. Em vez de fixar um único conjunto de hiperparâmetros e avaliar seu desempenho isoladamente, a meta-aprendizagem sistematiza essa avaliação em um espaço de configurações, extraíndo padrões que indicam quais escolhas são consistentemente mais eficazes em uma determinada classe de problemas.

Neste trabalho, essa abordagem foi aplicada ao pré-processamento dos sinais de EEG: em vez de uma configuração fixa, 75 combinações entre filtro, normalização

e janela temporal foram avaliadas, e os resultados de agrupamento de cada combinação foram consolidados para identificar quais configurações favorecem a discriminação das classes de IM na população estudada.

A implementação deste *pipeline* utiliza o MATLAB para o pré-processamento dos sinais e o agrupamento final, enquanto o Python realiza a extração de características por meio da *Time Series Feature Extraction Library* (TSFEL) e a seleção de características por meio da *Correlation-based Feature Selection* (CFS). A integração a partir da comunicação entre os dois ambientes ocorre por meio de arquivos intermediários; o MATLAB exporta os sinais pré-processados de cada sujeito e o Python retorna a matriz de características com os nomes das colunas embutidos, garantindo a rastreabilidade de cada descritor ao seu canal e tipo de origem.

3.3.1 Pré-processamento

O pré-processamento segue a mesma sequência do *pipeline* de referência, ou seja, é composto pelas etapas de filtragem, normalização e seleção temporal com a diferença de que cada uma dessas etapas não é definida em uma única configuração, mas avaliada em múltiplas variações. A combinação das variações disponíveis define uma grade de 75 configurações de pré-processamento, resultante do produto entre 5 opções de filtragem, 3 métodos de normalização e 5 janelas temporais, conforme detalhado na Tabela 4.

As 5 configurações de filtragem compreendem: ausência de filtro (sinal bruto), filtro passa-baixas em 30 Hz (configuração do *pipeline* de referência), passa-faixa de 8–30 Hz (banda um-beta completa), passa-faixa de 8–13 Hz (banda mu isolada) e passa-faixa de 13–30 Hz (banda beta isolada). Todos os filtros são do tipo Butterworth de 5ª ordem com filtragem bidirecional via *filtfilt()*. Para os filtros passa-faixa, a função *butter()* é chamada com o parâmetro *bandpass*, mantendo as características de fase zero.

As 3 opções de normalização são: ausência de normalização, *z-score* e Min-Max, que escalam cada canal individualmente ao longo das 3584 amostras do ensaio. Ambas as normalizações são aplicadas à janela completa de 7 segundos anterior à seleção temporal, de forma consistente com a decisão metodológica adotada na abordagem de referência.

As 5 janelas temporais avaliam diferentes segmentos do período pós-*onset*: janela tardia (2,0–4,5 s), janela de fim de tarefa (2,5–3,5 s), janela atual (0,5–2,5 s,

configuração de referência), janela média (0,5–3,5 s) e janela longa (0,5–4,5 s). Cada janela representa uma hipótese distinta sobre o intervalo mais informativo para a discriminação entres as classes de IM.

Tabela 4 - Diferentes opções de pré-processamento

| Filtro | Normalização | Janela temporal |
|------------------------|---------------------|------------------------|
| Sinal bruto | Não normalizado | 2,0 a 4,5 s |
| Passa-baixas (30 Hz) | | 2,5 a 3,5 s |
| Passa-faixa (8-30 Hz) | z-score | 0,5 a 2,5 s |
| Passa-faixa (8-13 Hz) | Min-Máx | 0,5 a 3,5 s |
| Passa-faixa (13-30 Hz) | | 0,5 a 4,5 s |

Fonte: Autoria própria (2026).

3.3.2 Extração de características

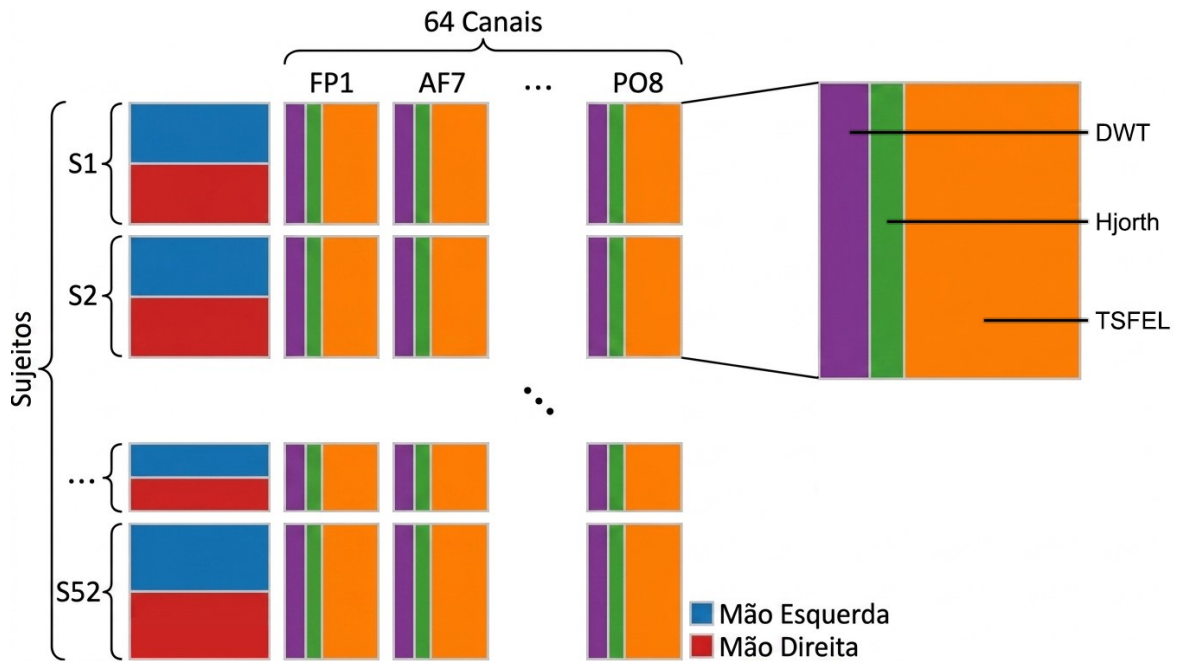
A extração de características combina as métricas DWT e os parâmetros de Hjorth descritos anteriormente com descritores adicionais extraídos pela TSFEL, uma biblioteca Python de código aberto que implementa um conjunto abrangente de descritores para séries temporais univariadas organizados em três domínios: temporal, espectral e estatístico. Para cada configuração da grade, o MATLAB exporta os sinais pré-processados de um sujeito em um arquivo intermediário que é lido pelo Python, no qual a TSFEL processa cada canal de cada ensaio e retorna as características extraídas.

A TSFEL extrai um total fixo de 156 características por canal, determinado pela configuração *get_features_by_domain()*, independentemente do tamanho da janela temporal. Os nomes das características espectrais que incluem valores de frequência refletem os bins da Transformada de Fourier de Tempo Curto calculada internamente pela biblioteca, cuja resolução é determinada pelo comprimento do sinal fornecido e pode variar entre diferentes configurações de janela.

Para garantir a rastreabilidade e viabilizar a comparação entre as 75 tabelas, as características são organizadas em uma tabela atributo-valor, em que cada coluna é identificada por um nome único no formato *Canal_Característica* (por exemplo,

C3_E_D4, FC4_mean, Cz_spectrogram_8.26hz), conforme ilustrado na Figura 8. Essa nomenclatura permite que o algoritmo de seleção opere pelo nome da coluna, e não por sua posição, tornando as tabelas comparáveis independentemente do pré-processamento. O Python salva a matriz de características e o vetor de nomes em um arquivo *.mat* que é lido pelo MATLAB, completando a ponte entre os dois ambientes.

Figura 8 - Representação estrutural da matriz de características extraídas de sinais de EEG.



Fonte: Autoria própria (2026).

3.3.3 Seleção de canais e características

Em substituição à PCA adotada no *pipeline* de referência, a redução de dimensionalidade foi realizada por meio da CFS, que seleciona subconjuntos de características com base em dois critérios simultâneos: alta correlação individual com a variável de classe (relevância) e baixa correlação mútua entre as próprias características (não redundância), conforme o critério de mérito apresentado na Equação 33.

$$Q_s = \frac{\kappa \bar{\rho}_{\phi\ell}}{\sqrt{\kappa + \kappa(\kappa - 1) \bar{\rho}_{\phi\phi}}} \quad (33)$$

onde Q_s representa o mérito do subconjunto S , κ é o número de características do subconjunto, $\bar{\rho}_{\phi\ell}$ corresponde à correlação média entre característica e classe, e $\bar{\rho}_{\phi\phi}$ representa a correlação média entre pares de características. A CFS foi implementada

em Python utilizando a biblioteca *scikit-feature*⁸, que disponibiliza o algoritmo de busca sequencial progressiva baseado no critério de mérito.

Dado o elevado número de características geradas pela combinação de DWT, Hjorth e TSFEL, a aplicação direta da CFS seria computacionalmente inviável, uma vez que o algoritmo apresenta complexidade proporcional ao número de características avaliadas. Para contornar essa limitação sem comprometer a qualidade da seleção, foi adotado um pré-filtro que reduz o espaço de busca aos 100 pares canal–característica com maior relevância discriminativa antes da execução da CFS.

Essa relevância foi estimada por meio da correlação absoluta de Pearson, calculada individualmente para cada sujeito, conforme a Equação 34. Essa estratégia foi adotada para evitar contaminação inter-indivíduo, uma vez que o cálculo da correlação sobre o conjunto global poderia favorecer artificialmente determinadas características em função de diferenças de amplitude entre os sujeitos. Assim, para cada sujeito, computou-se a correlação entre cada característica e o vetor de rótulos de classe, utilizando-se o valor absoluto do coeficiente como medida de relevância.

$$\rho = \frac{\sum_{i=1}^n (\phi_i - \bar{\phi})(\ell_i - \bar{\ell})}{\sqrt{\sum_{i=1}^n (\phi_i - \bar{\phi})^2} \sqrt{\sum_{i=1}^n (\ell_i - \bar{\ell})^2}}, \quad (34)$$

onde ϕ representa o vetor de valores de uma característica ao longo dos ensaios de um sujeito, ℓ corresponde ao vetor de rótulos de classe (0 para mão esquerda, 1 para mão direita) e ρ representa o coeficiente de correlação linear entre ambos. A relevância final de cada característica foi então obtida pela média aritmética dos valores absolutos dos 52 sujeitos, sendo selecionados os 100 pares canal–característica com maior relevância média.

O coeficiente de Pearson foi utilizado exclusivamente como mecanismo de redução do espaço de busca, preservando os pares canal–característica individualmente mais relevantes. Entretanto, a alta relevância individual não garante complementaridade entre características, uma vez que diferentes descritores podem conter informações redundantes. Dessa forma, os pares pré-selecionados foram submetidos à CFS, responsável por selecionar subconjuntos com elevada capacidade discriminativa conjunta e baixa redundância interna.

⁸ Disponível em: https://jundongl.github.io/scikit-feature/html/skfeature.function.statistical_based.CFS.html. Acesso em: 30 mai, 2026.

Como o *pipeline* considera 75 combinações distintas de pré-processamento, esse processo foi executado independentemente para cada configuração, resultando em 7.500 seleções. Cada seleção corresponde a um par canal–característica, permitindo avaliar simultaneamente a relevância espacial e a relevância descritiva dos atributos. Para identificar os elementos mais discriminativos de forma global, adotou-se um procedimento de votação por frequência de seleção, no qual cada ocorrência de um par canal–característica entre os subconjuntos selecionados pelo pré-filtro recebe um voto. Ao final, os votos são agregados separadamente para os 64 canais e as 163 características extraídas, de modo que aqueles com maior frequência de ocorrência entre as diferentes combinações sejam interpretados como os mais discriminativos e consistentes.

3.4 Agrupamento e avaliação de desempenho

A etapa final do processamento consistiu na identificação de padrões latentes nos dados por meio de aprendizado de máquina não supervisionado. A matriz reduzida F e a seleção de características da CFS serviram como entrada para o algoritmo de agrupamento, cujo objetivo foi particionar os ensaios em dois grupos distintos correspondentes às tarefas de imaginação motora, sem a utilização prévia dos rótulos de classe.

3.4.1 Algoritmo *k-means*

O agrupamento foi operacionalizado por meio da função `kmeans()`⁹, integrante da *Statistics and Machine Learning Toolbox* do MATLAB. O algoritmo particiona as observações iterativamente, minimizando a soma das distâncias quadráticas de cada ponto ao centroide do respectivo agrupamento. Foram definidos os seguintes parâmetros de configuração para a execução do algoritmo:

1. **Número de Clusters (K):** Fixado em 2, refletindo a estrutura binária de classes do problema experimental.
2. **Métrica de Distância:** Utilizou-se a distância euclidiana ao quadrado, adequada para espaços vetoriais contínuos onde as dimensões são ortogonais.
3. **Método de Inicialização:** Adotaram-se valores aleatórios para a seleção dos centroides iniciais.

⁹ Disponível em: <https://www.mathworks.com/help/stats/kmeans.html>. Acesso em: 3 dez, 2025

4. **Critério de Convergência:** O algoritmo foi configurado para convergir quando a variação relativa na soma das distâncias intra-*cluster* entre iterações consecutivas fosse inferior a 10^{-4} , ou quando o número máximo de 100 iterações por réplica fosse atingido.

3.4.2 Validação via índice de Rand ajustado

Após a aplicação do *k-means*, os rótulos de *cluster* gerados foram comparados aos rótulos reais das classes por meio do ARI, para avaliar a capacidade do modelo de discriminar corretamente os padrões de imaginação motora das duas tarefas. O cálculo do ARI foi realizado por meio da análise da matriz de contingência, obtida pela Tabela 2, entre os rótulos de *cluster* e os rótulos reais. Esta matriz, de dimensões $[K \times C]$, onde K é o número de *clusters* e C o número de classes, registra a frequência de observações que pertencem simultaneamente a cada combinação *cluster-classe*.

A partir desta matriz de contingência, o ARI foi computado considerando três aspectos fundamentais: (1) pares concordantes, que correspondem às observações agrupadas ou separadas de forma consistente em ambas as partições; (2) correção para o acaso, um ajuste baseado na concordância esperada para partições aleatórias com as mesmas distribuições marginais; e (3) normalização, garantindo que o índice varie no intervalo $[-1, 1]$ com interpretação padronizada. O algoritmo implementado seguiu a formulação matemática estabelecida no referencial teórico, processando a matriz de contingência para extrair os termos necessários ao cálculo do índice, com uma implementação modular que permite sua reutilização em diferentes contextos de validação de agrupamentos.

4 RESULTADOS E DISCUSSÕES

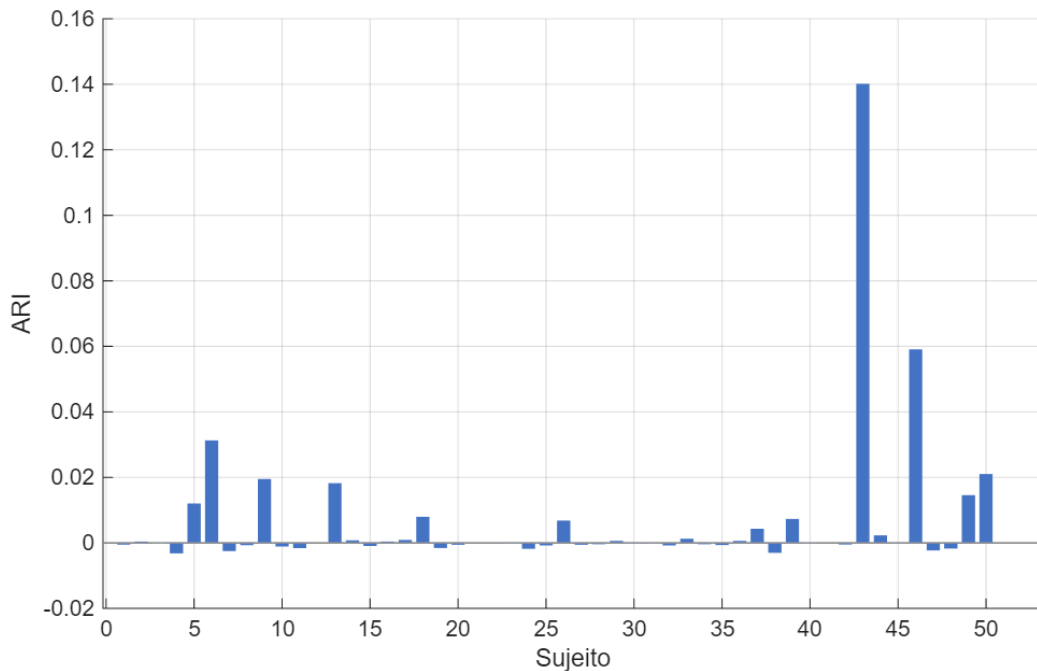
Este capítulo apresenta os resultados obtidos ao longo dos *pipelines* propostas, abrangendo as etapas de pré-processamento dos sinais de EEG, extração e seleção de características, e classificação das classes de IM. Dois *pipelines* são avaliados: o primeiro adota as configurações consolidadas na literatura como linha de base, enquanto o segundo investiga sistematicamente o impacto de diferentes combinações de filtros, janelas temporais e normalizações no desempenho dos classificadores. A comparação entre os dois *pipelines* constitui o eixo central da análise, permitindo identificar em que medida a exploração de configurações alternativas de pré-processamento é capaz de superar a abordagem de referência. Os resultados são analisados de forma agregada para o conjunto completo de 52 sujeitos, permitindo uma avaliação estatisticamente representativa do desempenho de cada configuração avaliada, com foco na identificação das combinações que melhor discriminam as classes de IM de mão esquerda e mão direita.

4.1 Pipeline de referência

Os resultados por sujeito, apresentados na Figura 9, revelam um desempenho predominantemente próximo de zero em toda a população. Do total de 52 sujeitos, 25 (48,1%) apresentaram ARI negativo, indicando agrupamentos piores do que o esperado ao acaso, e 6 (11,5%) obtiveram ARI exatamente igual a zero. Apenas 21 sujeitos (40,4%) apresentaram ARI positivo, dos quais somente 8 superaram o limiar de 0,01 e apenas 2 ultrapassaram 0,05, valores ainda distantes de qualquer utilidade prática para uma BCI.

O ARI médio obtido, considerando o número de PC necessários para atingir 95% da variância total, foi de 0,0062, o que reforça a hipótese de que o comportamento típico do *pipeline* é a ausência de estrutura discriminativa detectável. O sujeito s43 destacou-se como *outlier* positivo, com ARI de 0,1402, isto é, um valor expressivamente superior ao segundo melhor resultado, obtido por s46 com ARI de 0,0591. Em seguida, s06 (ARI = 0,0313), s50 (ARI = 0,0210) e s09 (ARI = 0,0195) compõem o grupo dos cinco melhores desempenhos. No extremo oposto, os sujeitos com pior desempenho foram s04 (ARI = -0,0032), s38 (ARI = -0,0031) e s07 (ARI = -0,0025), cujos valores negativos indicam que os agrupamentos gerados divergem sistematicamente da estrutura real dos dados.

Figura 9 - ARI por sujeito para o *pipeline* de referência.



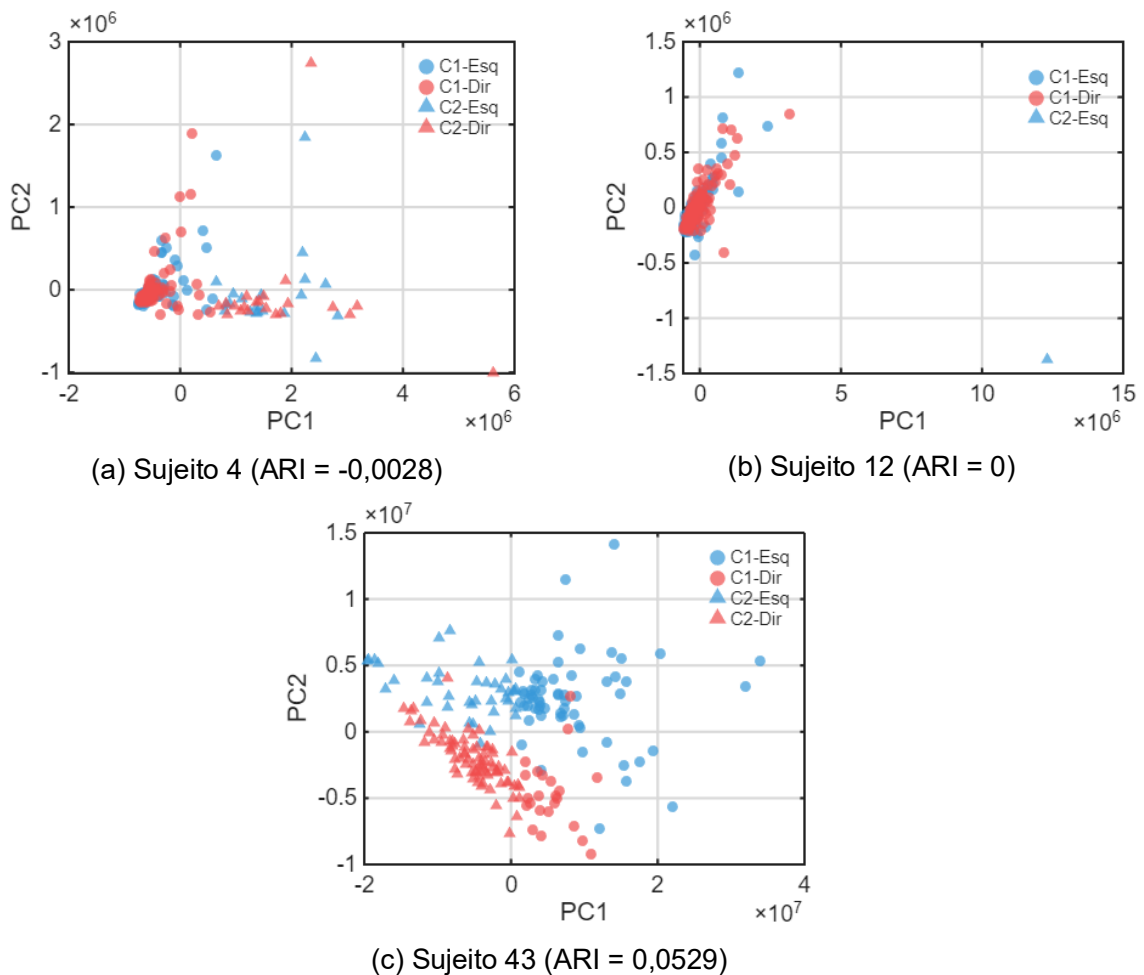
Fonte: Autoria própria (2026).

O desempenho modesto do *pipeline* de referência pode ser atribuído à atuação conjunta de dois fatores. O primeiro refere-se à restrição das características extraídas às sub-bandas D4 (16–32 Hz) e D5 (8–16 Hz), fundamentada no conhecimento clássico de ERD/ERS na IM, mas que pode desconsiderar contribuições discriminativas presentes em outras faixas de frequência. O segundo fator diz respeito à redução de dimensionalidade por PCA: embora o método preserve 95% da variância total, ele opera sobre as características já extraídas e não pode recuperar informação discriminativa que não esteja representada nas entradas. Dessa forma, se as características de entrada não capturam adequadamente a diferença entre as classes motoras, a PCA amplifica a estrutura presente nos dados sem introduzir nova capacidade discriminativa.

Na Figura 10(a), correspondente a s04, os pontos das duas classes encontram-se completamente sobrepostos e sem estrutura espacial aparente, com a maioria concentrada em uma região compacta e alguns pontos dispersos. O *k-means* divide o espaço de forma arbitrária, sem correspondência com os rótulos reais, o que explica o ARI negativo. Na Figura 10(b), referente a s12, os dados apresentam uma estrutura altamente colapsada ao longo do primeiro componente principal, sem separabilidade entre classes, resultando em ARI nulo, de modo que o agrupamento

não supera a de uma atribuição aleatória. Na Figura 10(c), s43 apresenta uma tendência de separação mais evidente entre as classes, estrutura que o *k-means* consegue capturar parcialmente, o que explica o maior ARI observado na população. Em todas as projeções, a legenda distingue os *clusters* gerados pelo *k-means*, representados por círculos (*Cluster 1*) e triângulos (*Cluster 2*), das classes reais de IM, indicadas por azul para a mão esquerda e por vermelho para mão direita, permitindo avaliar visualmente o grau de correspondência entre o agrupamento obtido e os rótulos verdadeiros.

Figura 10 - Visualização dos agrupamentos obtidos por *k-means* na projeção dos dois primeiros PC para diferentes sujeitos.

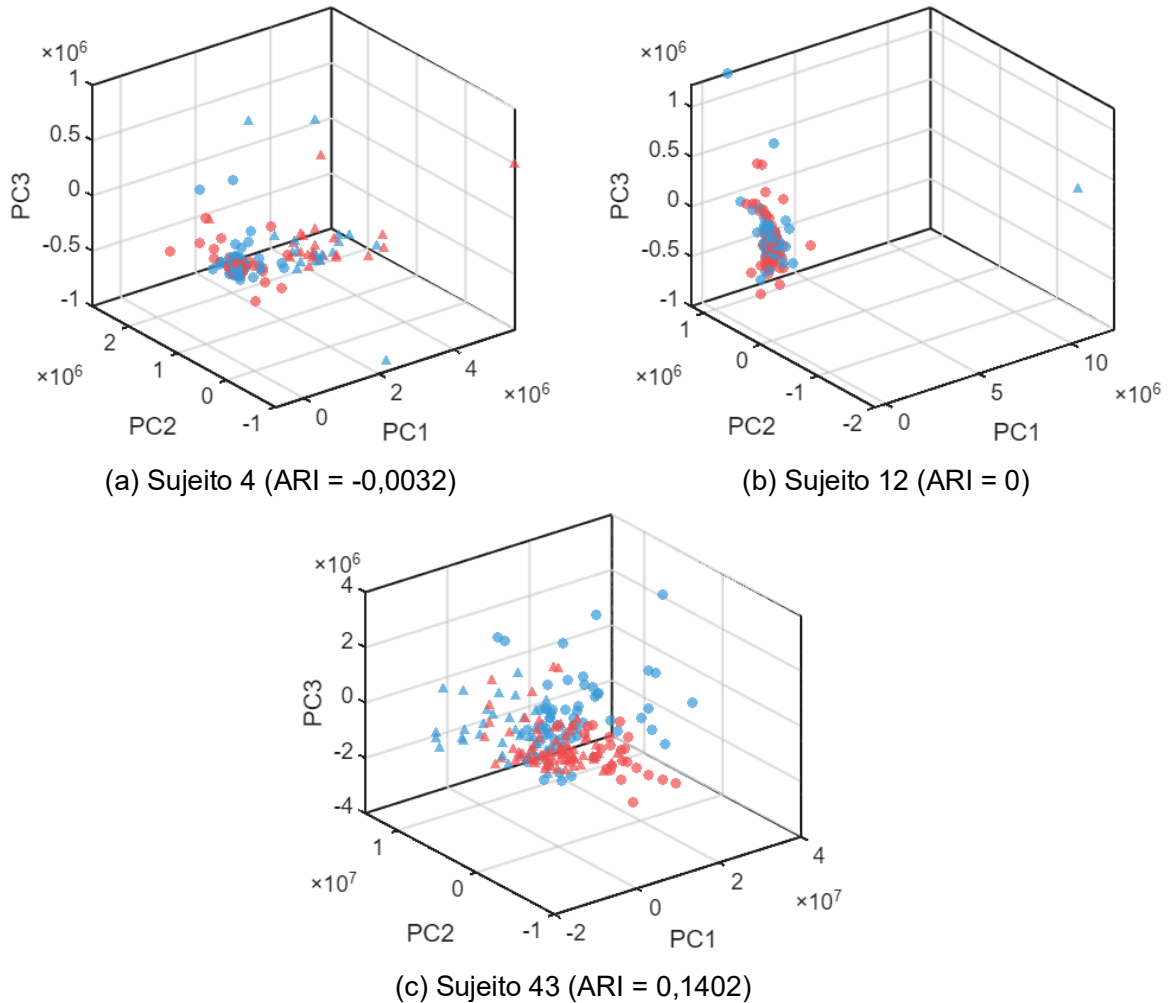


Fonte: Autoria própria (2026).

A projeção tridimensional complementa essa análise ao adicionar o terceiro componente principal. Na Figura 11(a) e na Figura 11(b), a inclusão do terceiro componente principal não revela estruturas adicionais para s04 e s12, de modo que os dados permanecem sobrepostos e sem separabilidade entre as classes. Para s43,

a Figura 11(c) reforça a tendência observada na projeção bidimensional, com os pontos de mão direita concentrados predominantemente na região inferior do espaço e os de mão esquerda na região superior, evidenciando que a separabilidade parcial desse sujeito se mantém consistente ao longo dos três primeiros PC.

Figura 11 - Visualização dos agrupamentos obtidos por *k-means* na projeção dos três primeiros PC para diferentes sujeitos.



Fonte: Autoria própria (2026).

A análise conjunta dos resultados numéricos e visuais indica que o *pipeline* de referência é capaz de capturar estrutura discriminativa apenas quando os dados do sujeito apresentam separabilidade natural no espaço dos PC, condição verificada em apenas uma minoria da população estudada, o que motiva a investigação de configurações alternativas de pré-processamento apresentadas na seção seguinte.

4.2 Pipeline de meta-aprendizagem

Os resultados do *pipeline* de meta-aprendizagem são apresentados nas Tabelas 5, 6 e 7, que exibem o ARI médio para todas as combinações de filtro e janela temporal sob as condições de dados não normalizados, normalização z-score e normalização min-máx, respectivamente. A Figura 12 sintetiza o impacto isolado de cada etapa do pré-processamento, permitindo uma análise comparativa.

A Figura 12(a) evidencia que o filtro é a etapa de maior influência no desempenho. O sinal não filtrado e o filtro passa-baixas de 30 Hz produziram ARI médio de 0,0989 e 0,1010, respectivamente, enquanto os filtros passa-faixa nas bandas 8–30 Hz, 8–13 Hz e 13–30 Hz resultaram em ARI médio de 0,0102, 0,0084 e 0,0066, respectivamente. A diferença entre os dois grupos é expressiva, com os filtros de banda ampla superando os passa-faixa em aproximadamente uma ordem de grandeza, indicando que a eliminação de componentes espectrais pelos filtros passa-faixa remove informação relevante para a separação das classes.

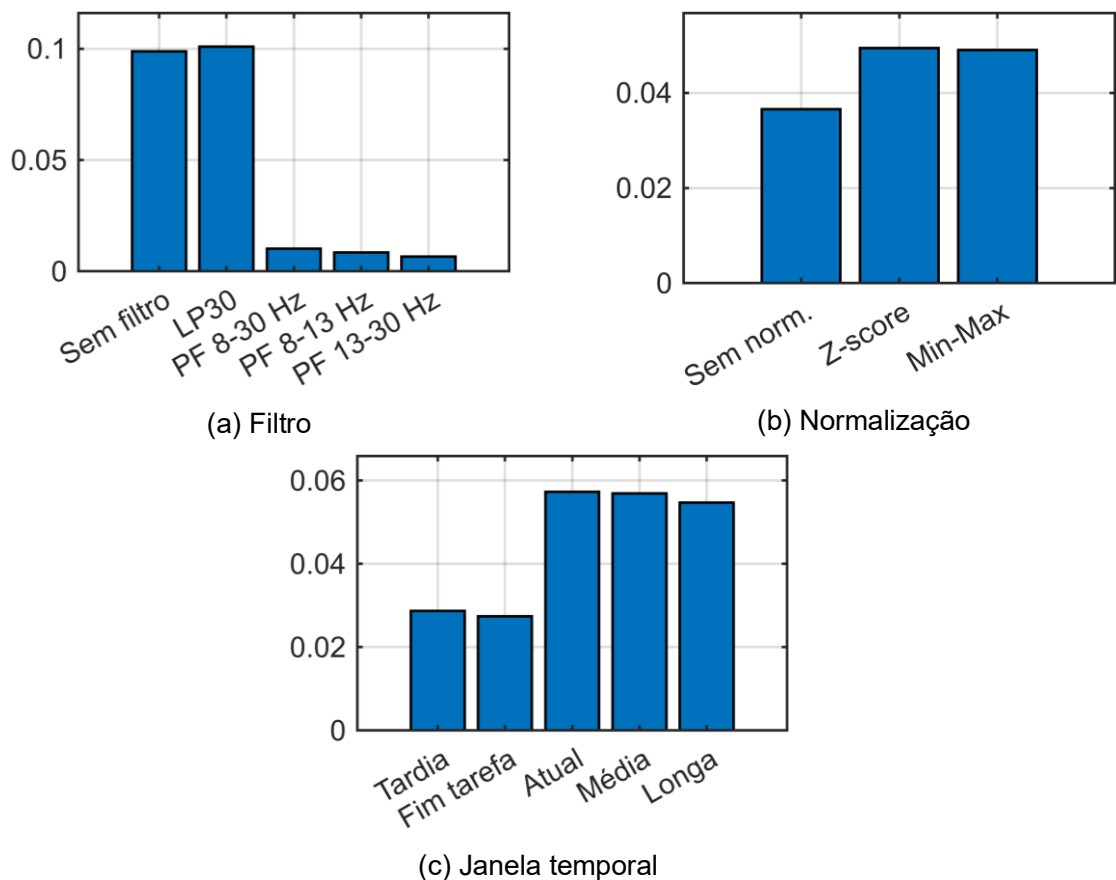
A superioridade dos filtros de banda ampla admite interpretações complementares. Do ponto de vista neurofisiológico, os filtros passa-faixa testados restringem o sinal às bandas mu e beta, descartando componentes de menor frequência que podem carregar informação discriminativa. A banda theta está bem estabelecida na literatura como marcador de planejamento motor, memória de trabalho e esforço mental sustentado. A banda delta, por sua vez, está associada a processos atencionais de alto nível e pode refletir aspectos do estado cognitivo do sujeito durante a tarefa. Ambas as bandas são preservadas pelo passa-baixas de 30 Hz, mas eliminadas pelos passa-faixa, o que pode explicar parte expressiva do ganho observado. Do ponto de vista técnico, contudo, não é possível descartar a contaminação por artefatos oculares de baixa frequência: movimentos oculares lentos ocorrem predominantemente na faixa delta e podem apresentar distribuição topográfica sistematicamente distinta entre ensaios de mão esquerda e mão direita, caso o olhar acompanhe a direção da ação imaginada. Como não foi aplicada remoção ativa de artefatos não é possível dissociar as contribuições neurofisiológicas das contaminações por artefatos nessa faixa de frequência.

A Figura 12(b) revela que a normalização z-score e min-máx produziram ARI médio de 0,0494 e 0,0490, respectivamente, enquanto os dados não normalizados resultaram em 0,0366. A diferença entre as duas normalizações é praticamente

inexistente, mas ambas superam a condição de dados não normalizados, indicando que o tratamento da escala das características favorece o processo de agrupamento, possivelmente por reduzir o domínio de características de maior magnitude que, nos dados não normalizados, poderiam enviesar a métrica de similaridade do *k-means*.

A Figura 12(c) demonstra um padrão claro de melhora à medida que aumenta a janela temporal. As janelas tardias e de fim de tarefa apresentaram ARI médio de 0,0287 e 0,0274, enquanto as janelas atual, média e longa convergiram para valores em torno de 0,057, representando um aumento de aproximadamente 108% em relação às janelas mais curtas. Esse resultado sugere que a inclusão de um intervalo temporal mais amplo captura padrões de IM mais estáveis e discriminativos, possivelmente por incorporar componentes de latência mais longa associadas ao processamento motor.

Figura 12 - Comparação do ARI médio por etapa do pré-processamento



Fonte: Autoria própria (2026).

A Tabela 5, referente à condição de dados não normalizados, evidencia que o sinal não filtrado e o filtro passa-baixas de 30 Hz produzem ARI médio

consistentemente superior aos filtros passa-faixa em todas as janelas temporais avaliadas. Os maiores valores são observados nas janelas atual (ARI = 0,1047 e 0,1189, respectivamente) e média (ARI = 0,1050 e 0,1039), enquanto os filtros passa-faixa não ultrapassam 0,0195 em nenhuma combinação. A janela tardia apresenta desempenho inferior às demais para os filtros de banda ampla, sugerindo que o intervalo temporal curto contém menor quantidade de informação discriminativa quando os dados não são normalizados.

Tabela 5 – ARI médio para as combinações de filtro e janela temporal sem normalização

| Filtro | Janela temporal | ARI médio |
|------------------------|-----------------|-----------|
| Sem filtro | Tardia | +0,0648 |
| | Fim de tarefa | +0,0072 |
| | Atual | +0,1047 |
| | Média | +0,1050 |
| | Longa | +0,1084 |
| Passa-baixas (30 Hz) | Tardia | +0,0593 |
| | Fim de tarefa | +0,0080 |
| | Atual | +0,1189 |
| | Média | +0,1039 |
| | Longa | +0,0977 |
| Passa-faixa (8-30 Hz) | Tardia | +0,0066 |
| | Fim de tarefa | +0,0058 |
| | Atual | +0,0137 |
| | Média | +0,0195 |
| | Longa | +0,0107 |
| Passa-faixa (8-13 Hz) | Tardia | +0,0040 |
| | Fim de tarefa | +0,0057 |
| | Atual | +0,0118 |
| | Média | +0,0110 |
| | Longa | +0,0085 |
| Passa-faixa (13-30 Hz) | Tardia | +0,0028 |
| | Fim de tarefa | +0,0042 |
| | Atual | +0,0111 |
| | Média | +0,0115 |
| | Longa | +0,0104 |

Fonte: Autoria própria (2026).

A Tabela 6, correspondente à normalização *z-score*, mantém o padrão observado na condição anterior, mas apresenta valores de ARI superiores para os filtros de banda ampla. Vale destacar que a combinação de filtro passa-baixas de 30 Hz, normalização *z-score* e janela atual corresponde exatamente às etapas de pré-processamento adotadas no *pipeline* de referência, e os resultados obtidos nessa configuração (ARI = 0,1336) evidenciam que, mesmo sob condições idênticas de pré-processamento, o *pipeline* de meta-aprendizagem supera expressivamente o *pipeline* de referência, cujo ARI médio foi de 0,0062. O maior ARI da tabela é obtido pela combinação do sinal não filtrado com janela média (ARI = 0,1432), seguido pelo filtro

passa-baixas de 30 Hz com janela longa (ARI = 0,1364) e pela janela atual. Destaca-se, ainda, que a normalização z-score promove um ganho expressivo na janela de fim de tarefa para os filtros de banda ampla, que passa de valores próximos a 0,007 na Tabela 5 para valores acima de 0,101, indicando que essa janela temporal é particularmente sensível à escala dos dados.

Tabela 6 – ARI médio para as combinações de filtro e janela temporal utilizando normalização z-score

| Filtro | Janela temporal | ARI médio |
|------------------------|------------------------|------------------|
| Sem filtro | Tardia | +0,0519 |
| | Fim de tarefa | +0,1016 |
| | Atual | +0,1324 |
| | Média | +0,1432 |
| | Longa | +0,1352 |
| Passa-baixas (30 Hz) | Tardia | +0,0509 |
| | Fim de tarefa | +0,1015 |
| | Atual | +0,1336 |
| | Média | +0,1420 |
| | Longa | +0,1364 |
| Passa-faixa (8-30 Hz) | Tardia | +0,0048 |
| | Fim de tarefa | +0,0061 |
| | Atual | +0,0085 |
| | Média | +0,0125 |
| | Longa | +0,0098 |
| Passa-faixa (8-13 Hz) | Tardia | +0,0051 |
| | Fim de tarefa | +0,0042 |
| | Atual | +0,0085 |
| | Média | +0,0104 |
| | Longa | +0,0082 |
| Passa-faixa (13-30 Hz) | Tardia | +0,0026 |
| | Fim de tarefa | +0,0017 |
| | Atual | +0,0042 |
| | Média | +0,0142 |
| | Longa | +0,0061 |

Fonte: Autoria própria (2026).

A Tabela 7, referente à normalização min-máx, apresenta os maiores valores absolutos de ARI entre as três condições, com destaque para o filtro passa-baixas de 30 Hz, com janela atual (ARI = 0,1454) e janela longa (ARI = 0,1426), além do sinal não filtrado, com janela atual (ARI = 0,1359). O comportamento da janela de fim de tarefa permanece elevado para os filtros de banda ampla, corroborando a observação de que a normalização beneficia especialmente essa configuração temporal. Os filtros passa-faixa, independentemente da normalização aplicada, não ultrapassam ARI de 0,0209 em nenhuma combinação avaliada.

Tabela 7 – ARI médio para as combinações de filtro e janela temporal utilizando normalização min-máx

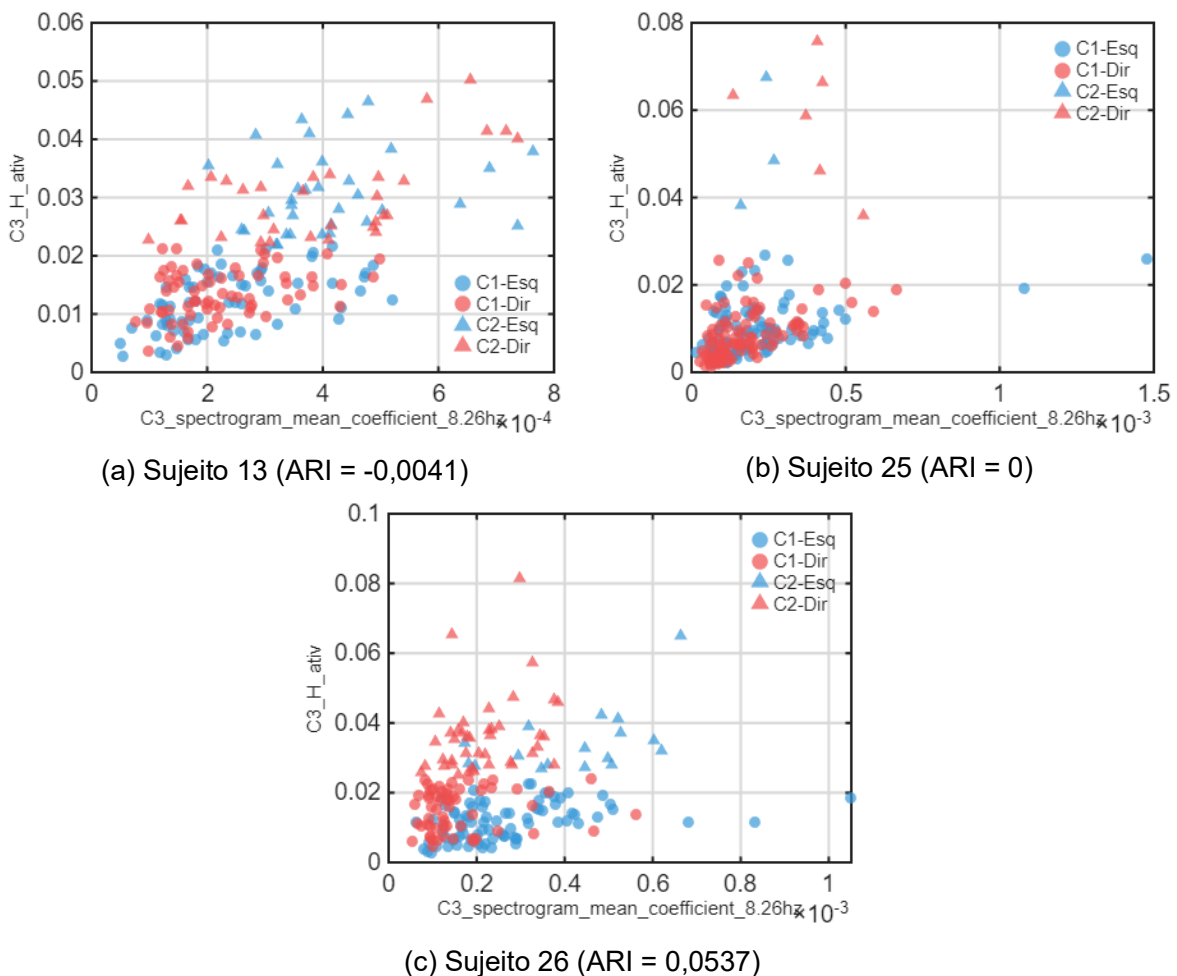
| Filtro | Janela temporal | ARI médio |
|------------------------|------------------------|------------------|
| Sem filtro | Tardia | +0,0795 |
| | Fim de tarefa | +0,0756 |
| | Atual | +0,1359 |
| | Média | +0,1233 |
| | Longa | +0,1146 |
| Passa-baixas (30 Hz) | Tardia | +0,0855 |
| | Fim de tarefa | +0,0755 |
| | Atual | +0,1454 |
| | Média | +0,1133 |
| | Longa | +0,1426 |
| Passa-faixa (8-30 Hz) | Tardia | +0,0049 |
| | Fim de tarefa | +0,0056 |
| | Atual | +0,0132 |
| | Média | +0,0209 |
| | Longa | +0,0098 |
| Passa-faixa (8-13 Hz) | Tardia | +0,0072 |
| | Fim de tarefa | +0,0075 |
| | Atual | +0,0090 |
| | Média | +0,0111 |
| | Longa | +0,0131 |
| Passa-faixa (13-30 Hz) | Tardia | +0,0012 |
| | Fim de tarefa | +0,0012 |
| | Atual | +0,0082 |
| | Média | +0,0115 |
| | Longa | +0,0083 |

Fonte: Autoria própria (2026).

Após a aplicação da CFS a cada uma das 75 combinações, as características selecionadas foram submetidas a uma votação por frequência de seleção, identificando os pares canal-característica mais recorrentes ao longo das configurações avaliadas. As três características mais frequentemente selecionadas foram *C3_spectrogram_mean_coeficient_8.26hz*, que representa o coeficiente médio do espectrograma na frequência de 8,26 Hz e está relacionado à banda alfa/mu associada à supressão do ritmo mu durante a IM, *C3_H_ativ*, correspondente ao parâmetro de Atividade de Hjorth que quantifica a variância do sinal e reflete a potência total do EEG no canal, e *C3_mean_absolute_deviation*, que expressa o desvio absoluto médio do sinal no domínio do tempo e constitui uma medida de dispersão robusta a valores atípicos. Todas as três características provêm do canal C3, localizado sobre o córtex motor primário do hemisfério esquerdo. Esse resultado é neurologicamente coerente com a tarefa de IM bilateral, uma vez que C3 está associado ao controle e à imaginação de movimentos da mão direita e figura como referência consolidada na literatura de BCI.

A Figura 13 apresenta a visualização bidimensional dos agrupamentos para três sujeitos representativos, utilizando dois pares de canal-característica com maior frequência de seleção. Na Figura 13(a), correspondente a s13 (ARI = -0,0041), os pontos das duas classes encontram-se completamente sobrepostos, sem estrutura discriminativa aparente. Na Figura 13(b), referente a s25 (ARI = 0), os dados concentram-se em uma faixa estreita ao longo do eixo horizontal, sem separabilidade entre as classes. Na Figura 13(c), s26 (ARI = 0,0537) apresenta uma estrutura mais favorável, com os triângulos do *cluster* C2 concentrados em valores mais elevados de $C3_H_ativ$ e os círculos do *cluster* C1 distribuídos em valores mais baixos, permitindo ao *k-means* uma separação parcial entre as classes.

Figura 13 - Visualização bidimensional dos agrupamentos utilizando os pares canal-característica mais frequentemente selecionados.

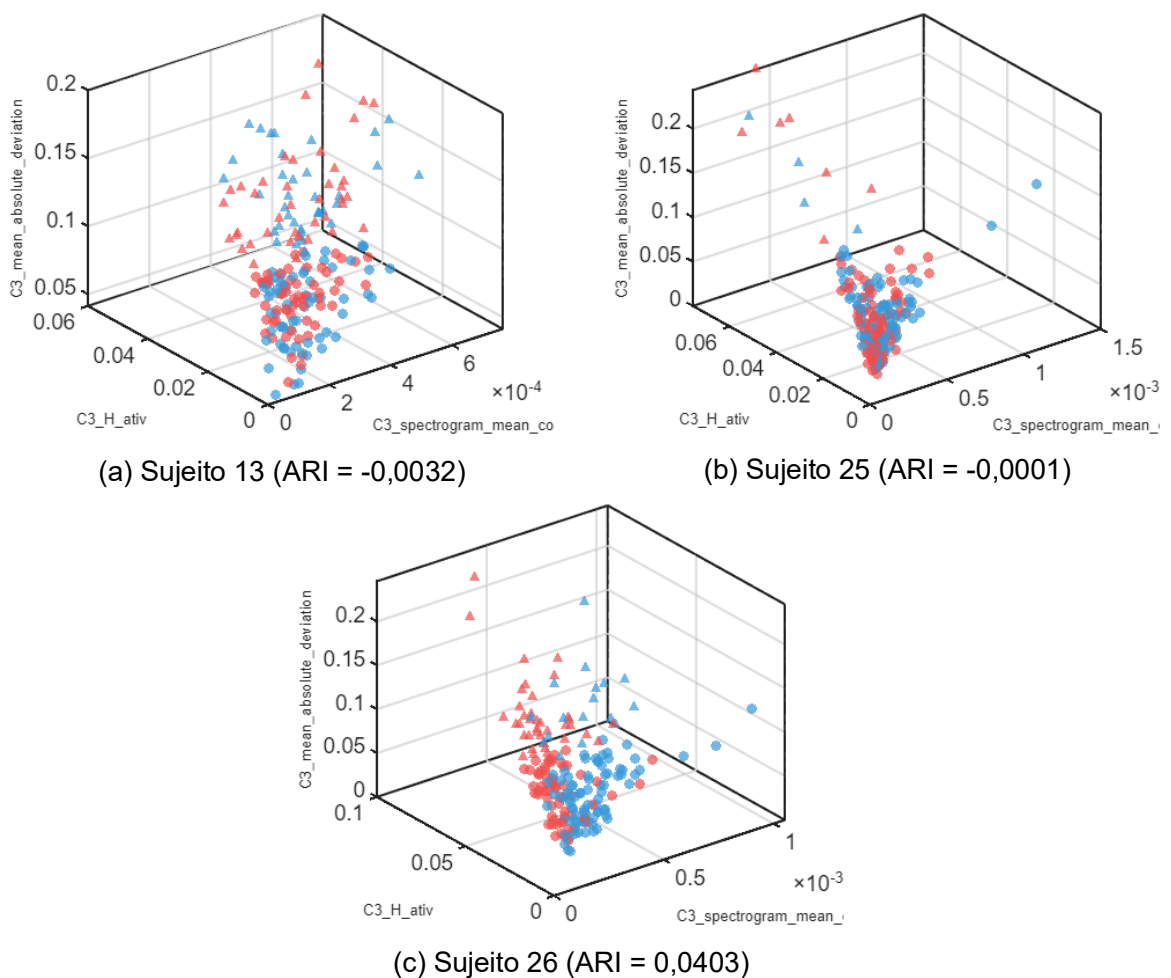


Fonte: Autoria própria (2026).

A Figura 14 adiciona o terceiro par de canal-característica mais selecionado, $C3_mean_absolute_deviation$, à projeção tridimensional. Na Figura 14(a) e na Figura

14(b), a dimensão adicional não revela estrutura discriminativa para s13 e s25, com os dados permanecendo sobrepostos ao longo dos três eixos. Para s26, a Figura 14(c) reforça a separabilidade observada na projeção bidimensional, com ARI de 0,0403, ainda que a sobreposição entre as classes permaneça considerável. A concentração das características mais selecionadas no canal C3 e sua interpretabilidade neurofisiológica reforçam a coerência do processo de seleção adotado.

Figura 14 - Visualização tridimensional dos agrupamentos utilizando os pares canal-característica mais frequentemente selecionados.



Fonte: Autoria própria (2026).

4.3 Análise das características selecionadas

A análise dos pares canal-característica, sintetizada na Tabela 8, revela que os dez pares mais frequentemente selecionados são todos provenientes do canal C3, evidenciando uma concentração notável da informação discriminativa nessa região. O par mais recorrente combina C3 com o coeficiente espectral na frequência de 8,26 Hz, presente em 45,3% das combinações avaliadas, o que reforça o papel do ritmo

mu na modulação do sinal durante a IM. Os demais pares do top 10 alternam entre medidas de dispersão no domínio do tempo, como variância, desvio padrão e desvio absoluto médio, e coeficientes *wavelet* na banda beta inferior (14 a 19 Hz), sugerindo que a discriminação entre as classes se manifesta de forma complementar nos domínios temporal e espectral. A dominância exclusiva de C3 no topo do ranqueamento, com o primeiro par proveniente de outro canal surgindo apenas na 18ª posição, reafirma que a informação discriminativa está fortemente localizada no córtex motor primário do hemisfério esquerdo, independentemente da combinação.

Após a aplicação da CFS a cada uma das 75 combinações de pré-processamento, cada combinação gerou um conjunto de 100 pares canal-característica, inicialmente selecionados por correlação de Pearson. Vale destacar que a CFS retornou o mesmo conjunto em todos os casos, confirmando que os 100 pares selecionados pela correlação de Pearson já satisfazem o critério de relevância e baixa redundância da CFS, o que indica coerência do processo de seleção adotado. Para identificar os canais e características mais relevantes de forma global, os votos de cada par canal-característica foram acumulados ao longo das 75 combinações, submetidos à transformação logarítmica na base 10 e normalizados pelo método min-máx, gerando a frequência normalizada de seleção exibida na Figura 15.

Tabela 8 - Top 10 pares canal-característica por frequência de seleção.

| Posição | Canal-Characterística | Nº de votos |
|---------|--|-------------|
| 1 | C3_spectrogram_mean_coefficient_8.26hz | 34 |
| 2 | C3_H_ativ | 31 |
| 3 | C3_mean_absolute_deviation | 31 |
| 4 | C3_variance | 31 |
| 5 | C3_standard_deviation | 30 |
| 6 | C3_wavelet_variance_18.29hz | 29 |
| 7 | C3_wavelet_energy_18.29hz | 29 |
| 8 | C3_wavelet_standard_deviation_18.29hz | 29 |
| 9 | C3_wavelet_standard_deviation_16.0hz | 29 |
| 10 | C3_wavelet_energy_16.0hz | 29 |

Fonte: Autoria própria (2026).

O mapa topográfico da Figura 15 revela uma concentração expressiva nas regiões centrais esquerdas e frontotemporais direitas do escalpo. Para interpretar os resultados, é importante distinguir duas métricas complementares utilizadas na análise: o número de votos, que corresponde ao total de vezes que características de um determinado canal foram selecionadas pelo CFS ao longo das 75 combinações (denominador de 7.500, resultante de 75 combinações × 100 características cada); e

a presença, que indica em quantas das 75 combinações aquele canal contribuiu com ao menos uma característica selecionada.

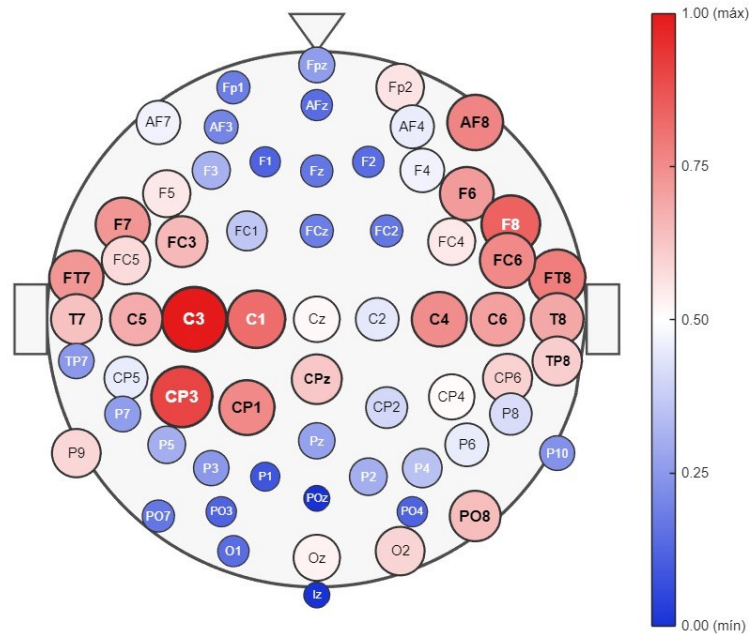
C3 foi o canal mais discriminativo segundo ambas as métricas: esteve presente em todas as 75 combinações avaliadas e acumulou 1.166 votos, correspondendo a 15,5% do total. Esse resultado destaca C3 de forma expressiva em relação aos demais canais. Em seguida, CP3 acumulou 660 votos (8,8% do total), F8 com 457 votos (6,1%), C1 com 402 votos (5,4%) e FT8 com 326 votos (4,3%), compondo os cinco canais mais frequentemente selecionados. A predominância de C3 e CP3 é neurofisiologicamente coerente com a tarefa de IM bilateral: ambos os canais se localizam sobre o córtex motor primário e a região parietal superior do hemisfério esquerdo, regiões classicamente associadas ao planejamento e à execução de movimentos da mão direita, possivelmente consequência da predominância de indivíduos destros no banco de dados. A presença expressiva de F8, FT8 e AF8 aponta para uma contribuição relevante do hemisfério direito, como esperado em uma tarefa que envolve a imaginação de movimentos bilaterais. Em contrapartida, os canais das regiões occipital e parietal medial apresentaram frequência de seleção próxima ao mínimo, o que é consistente com a natureza motora da tarefa, sem envolvimento primário do córtex visual.

A análise por janela temporal revela que C3 e CP3 mantiveram presença consistente em todas as cinco janelas. C1 apresentou um comportamento distinto, com baixa frequência nas janelas tardias e de fim de tarefa e aumento expressivo nas demais janelas, sugerindo que sua contribuição discriminativa está associada a janelas temporais mais amplas. C4, canal simétrico a C3 no hemisfério direito, apresentou frequência relevante nas três primeiras janelas, mas ausência nas janelas média e longa, indicando que sua relevância é mais restrita a intervalos temporais curtos ao se tratar de indivíduos majoritariamente destros.

Entretanto, é necessário considerar a composição demográfica do banco de dados: dos 52 sujeitos, 50 são destros e apenas 2 são ambidestros, sem a presença de sujeitos canhotos. Essa distribuição assimétrica impede dissociar, com os dados disponíveis, se a dominância de C3 reflete uma propriedade geral da IM bilateral ou se é, ao menos parcialmente, uma assinatura neural específica da população destra, na qual o hemisfério esquerdo é sistematicamente dominante para o controle motor. Em um conjunto de dados balanceado entre destros e canhotos, é esperado que sujeitos canhotos demonstrem dominância do canal C4. A ausência de sujeitos

canhotos na amostra impede essa verificação e constitui uma limitação para a generalização dos resultados a populações com perfis de lateralidade distintos.

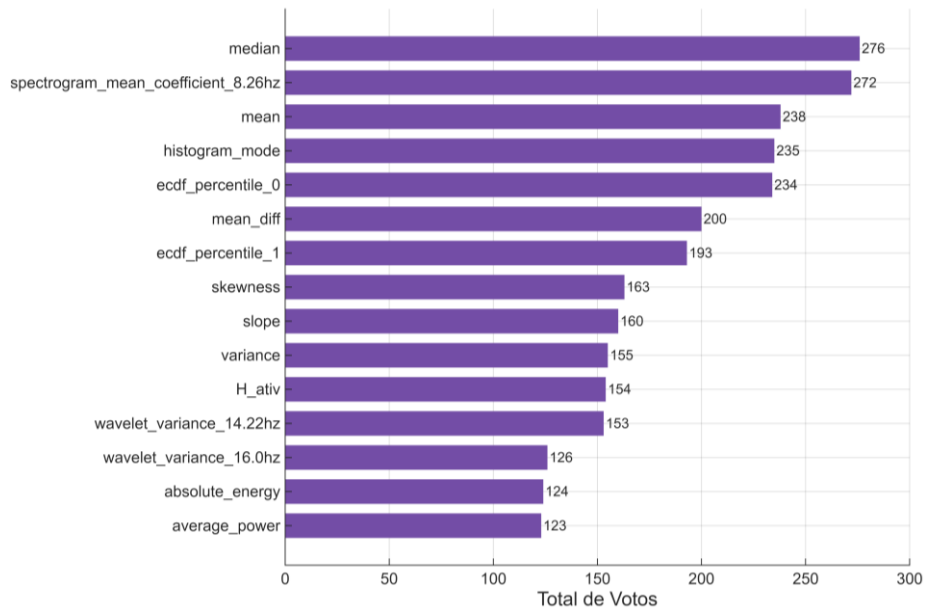
Figura 15 - Frequência normalizada de seleção dos canais do EEG nas combinações avaliadas.



Fonte: Autoria própria (2026).

De modo semelhante à obtenção dos canais, realizou-se a contagem das características, conforme ilustrado na Figura 16. A mediana foi a característica mais frequentemente selecionada ao longo das 75 combinações, seguida pelo coeficiente médio do espectrograma na frequência de 8,26 Hz, pela média, pelo modo do histograma e pelo percentil 0 da função de distribuição empírica acumulada. A recorrência de medidas estatísticas de tendência central e de distribuição, como mediana, média e modo do histograma, sugere que a separação entre as classes de IM se manifesta predominantemente por deslocamentos na distribuição das amplitudes do sinal. A presença do coeficiente espectral em 8,26 Hz entre as características mais selecionadas é particularmente relevante, pois essa frequência situa-se na banda alfa/mu, cujo ritmo é suprimido durante a ERD no processo de IM. O parâmetro de Atividade de Hjorth, que quantifica a variância do sinal e reflete sua potência total, também figura entre as características mais relevantes, reforçando que alterações na potência do sinal em C3 constituem um marcador discriminativo consistente para a tarefa avaliada.

Figura 16 - Frequência de seleção das características do EEG nas combinações avaliadas.



Fonte: Autoria própria (2026).

É importante ressaltar que o *pipeline* de meta-aprendizagem proposto neste trabalho é integralmente *offline* e exploratório. A extração de 156 características por canal por meio da TSFEL, aplicada a todas as 75 combinações de pré-processamento e a todos os 52 sujeitos, implica custo computacional elevado que inviabilizaria sua execução em tempo real em hardware de baixo custo, como aquele disponível em contextos clínicos de menor infraestrutura. Entretanto, essa limitação não compromete a aplicabilidade prática dos resultados: a fase de meta-aprendizagem seria conduzida *offline*, em hardware convencional, com o objetivo de identificar quais características e configurações de pré-processamento são mais discriminativas para uma determinada população. Uma vez estabelecido esse subconjunto apenas essas características selecionadas precisariam ser calculadas em tempo real no sistema, resultando em uma carga computacional substancialmente reduzida. Nesse sentido, a abordagem proposta pode ser interpretada como uma fase de calibração *offline* que habilita um sistema *online* eficiente, compatível com as restrições de hardware presentes em cenários de aplicação prática.

5 CONSIDERAÇÕES FINAIS

Este trabalho apresentou o desenvolvimento e a validação de um procedimento metodológico para a análise não supervisionada de sinais de EEG em tarefas de IM, com o objetivo de investigar a estrutura latente dos dados e sua correspondência às classes experimentais. Dois *pipelines* foram propostas e comparadas: uma de referência, fundamentada em configurações consolidadas na literatura, e outra de meta-aprendizagem, que explorou sistematicamente 75 combinações de filtros, janelas temporais e normalizações, combinadas com extração de características, seleção por correlação de Pearson e CFS, e agrupamento por *k-means*.

Os resultados demonstraram que o *pipeline* de meta-aprendizagem superou expressivamente o *pipeline* de referência em todas as métricas avaliadas, com a melhor combinação alcançando ARI de 0,1454 frente ao ARI médio de 0,0062 do *pipeline* de referência. A análise do impacto das etapas de pré-processamento revelou que o filtro é o fator de maior influência sobre o desempenho, com os filtros de banda ampla superando os passa-faixa em aproximadamente uma ordem de grandeza, e que janelas temporais mais longas capturam padrões de IM mais discriminativos. A análise das características selecionadas evidenciou a dominância do canal C3 em todas as combinações avaliadas, resultado neurologicamente coerente com a natureza bilateral da tarefa e consistente com a literatura de BCI. A concentração das features mais selecionadas em medidas estatísticas de distribuição e no coeficiente espectral na banda alfa/mu reforça que a discriminação entre as classes se manifesta tanto no domínio temporal quanto no espectral.

Os resultados obtidos são promissores e sugerem que a metodologia proposta contribui para a compreensão dos mecanismos neurais subjacentes à IM e à variabilidade de desempenho em interfaces cérebro-computador. A identificação de configurações de pré-processamento e de pares canal-característica com maior poder discriminativo constitui uma base sólida para o desenvolvimento de sistemas de BCI mais adaptativos e personalizados, capazes de lidar com a alta variabilidade intersujeito característica dos sinais de EEG.

Para superar os resultados obtidos, em trabalhos futuros, é necessária a expansão da metodologia analítica para o domínio do aprendizado supervisionado, utilizando os rótulos previamente conhecidos das tarefas para treinar modelos

predictivos robustos. Nesse cenário, a aplicação da análise discriminante linear destaca-se como um passo fundamental, dada a sua consagração na literatura de BCI para a maximização da separabilidade espacial das classes e sua alta eficiência computacional, o que viabiliza futuras aplicações em tempo real. Adicionalmente, no âmbito da aprendizagem não supervisionada, recomenda-se a substituição do algoritmo *k-means* por abordagens mais resilientes à alta dimensionalidade e tolerantes a ruídos de biopotenciais, como o DBSCAN que é fundamentado em densidade local, de forma a identificar regiões de alta densidade de pontos separadas por regiões esparsas e classificar pontos isolados como ruído, em vez de forçá-los a pertencer a um dos agrupamentos definidos a priori. Para dados de EEG, essa capacidade de tratar ensaios atípicos como ruído é metodologicamente vantajosa, pois evita que a posição dos centroides seja distorcida por observações espúrias.

Em termos de modelagem avançada de padrões, a aplicação de técnicas de aprendizado profundo, como redes neurais convolucionais e arquiteturas baseadas em *transformers*, representa uma direção promissora para capturar dependências temporais e topológicas mais complexas nos sinais de EEG. Por fim, a investigação de métodos de transferência de aprendizado, aliada à incorporação de *pipelines* de seleção de características adaptativos por indivíduo, apresenta-se como uma estratégia essencial para mitigar a inerente variabilidade intersujeito, aprimorando de forma significativa a discriminabilidade e a generalização dos sistemas de BCI populacionais

REFERÊNCIAS

- ACHARYA, Jayant N. *et al.* American Clinical Neurophysiology Society Guideline 2: Guidelines for Standard Electrode Position Nomenclature. **Journal of Clinical Neurophysiology**, v. 33, n. 4, p. 308–311, ago. 2016.
- ADUR, R. Sistema de processamento de sinais biomédicos: módulo didático de eletroencefalograma. 2008.
- AHMED, M.; SERAJ, R.; ISLAM, S. M. S. The k-means Algorithm: A Comprehensive Survey and Performance Evaluation. **Electronics**, v. 9, n. 8, p. 1295, 12 ago. 2020.
- AKBULUT, O. Feature Normalization Effect in Emotion Classification based on EEG Signals. **Sakarya University Journal of Science**, v. 24, n. 1, p. 60–66, 1 fev. 2020.
- ANDRADE, M. K. D. S. Detecção e classificação de imagética motora utilizando sinais de EEG e aprendizado de máquina. 2019.
- ANGHINAH, R. *et al.* Artefatos biológicos no EEG quantitativo. **Arquivos de Neuro-Psiquiatria**, v. 64, n. 2a, p. 264–268, jun. 2006.
- ASANZA, V.; PELAEZ, E.; LOAYZA, F. EEG signal clustering for motor and imaginary motor tasks on hands and feet. *In: IEEE*, out. 2017.
- BRAND, L.; SANTOS, P. S.; MIDORIKAWA, T. Y. **Aquisição e condicionamento de sinais cerebrais por eletroencefalografia.** [S.l.: S.n.].
- CALDAS, A. S. C. *et al.* Motor imagery and swallowing: a systematic literature review. **Revista CEFAC**, v. 20, n. 2, p. 247–257, abr. 2018.
- CHEN, D. *et al.* A high-performance seizure detection algorithm based on Discrete Wavelet Transform (DWT) and EEG. **PLOS ONE**, v. 12, n. 3, p. e0173138, 9 mar. 2017.
- CHO, Hohyun *et al.* **EEG datasets for motor imagery brain-computer interface.** GigaScienceOxford University Press, , 1 jul. 2017.
- DE ALMEIDA, W. F.; LIMA, C. A. D. M.; PERES, S. M. A systematic mapping of feature extraction and feature selection methods of electroencefalogram signals for neurological diseases diagnostic assistance. **IEEE Latin America Transactions**, v. 19, n. 5, p. 735–745, maio 2021.
- DE ASSIS, E. C. **Algoritmos de particionamento aplicados à análise estatística de formas.** [S.l.: S.n.].
- DECETY, J. The neurophysiological basis of motor imagery. **Behavioural Brain Research**, v. 77, n. 1–2, p. 45–52, 1 maio 1996.

DINIZ, P. S. R.; DA SILVA, E. A. B.; NETTO, S. L. **Digital signal processing: system analysis and design**. [S.l.]: Cambridge University Press, 2010.

FREEMAN, W. J.; QUIROGA, R. Q. **Imaging brain function with EEG: Advanced temporal and spatial analysis of electroencephalographic signals**. [S.l.]: Springer New York, 2013. v. 9781461449843

GANDHI, T.; PANIGRAHI, B. K.; ANAND, S. A comparative study of wavelet families for EEG signal classification. **Neurocomputing**, v. 74, n. 17, p. 3051–3057, out. 2011.

GÉRON, Aurélien. **Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems**. [S.l.]: O'Reilly Media, Inc., 2019.

GOMES, I. **Pessoas com deficiência têm menor acesso à educação, ao trabalho e à renda**. Disponível em: <https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/37317-pessoas-com-deficiencia-tem-menor-acesso-a-educacao-ao-trabalho-e-a-renda>. Acesso em: 21 jun. 2025.

HAZARIKA, N. *et al.* Classification of EEG signals using the wavelet transform. **Signal Processing**, v. 59, n. 1, p. 61–72, 1 maio 1997.

HJORTH, Bo. EEG analysis based on time domain properties. **Electroencephalography and Clinical Neurophysiology**, v. 29, n. 3, p. 306–310, 1 set. 1970.

HÖLLER, Yvonne *et al.* Comparison of EEG-Features and Classification Methods for Motor Imagery in Patients with Disorders of Consciousness. **PLoS ONE**, v. 8, n. 11, p. e80479, 25 nov. 2013.

HUANG, Z.; WANG, M. A review of electroencephalogram signal processing methods for brain-controlled robots. **Cognitive Robotics**, v. 1, p. 111–124, 2021.

IZBICKI, R.; DOS SANTOS, T. M. **Aprendizado de máquina: uma abordagem estatística**. [S.l.]: Rafael Izbicki, 2020.

JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data clustering. **ACM Computing Surveys**, v. 31, n. 3, p. 264–323, set. 1999.

JIMAN, A. A.; ATTAR, E. T. Unsupervised clustering of pre-ictal EEG in children: a reproducible and lightweight CPU-based workflow. **Frontiers in Neurology**, v. 16, 22 out. 2025.

JOLLIFFE, I. T. .. **Principal component analysis**. [S.l.]: Springer, 2002.

JOLLIFFE, I. T.; CADIMA, J. Principal component analysis: a review and recent developments. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 374, n. 2065, p. 20150202, 13 abr. 2016.

KORATS, Gundars *et al.* Applying ICA in EEG: Choice of the Window Length and of the Decorrelation Method. *In: [S.l.: S.n.]*. p. 269–286.

LAGE, A. C. Análise de novos dados linguísticos: a eletroencefalografia em neurociência da linguagem. **Revista FSA**, v. 10, n. 1, p. 153–172, 1 fev. 2013.

LAMEIRA, A. P. *et al.* Postura da mão e imagética motora: um estudo sobre reconhecimento de partes do corpo. **Revista Brasileira de Fisioterapia**, v. 12, n. 5, p. 379–385, out. 2008.

LOTZE, M.; HALSBAND, U. Motor imagery. **Journal of Physiology-Paris**, v. 99, n. 4–6, p. 386–395, 1 jun. 2006.

MELO JUNIOR, G. D. *et al.* Pré-processamento de sinais Eletro-oculográficos (EOG). **Research, Society and Development**, v. 10, n. 4, p. e28410414221, 10 abr. 2021.

NASCIMENTO, I. A. P. D. S. **Efeitos do treinamento da imagética motora da marcha em indivíduos com doença de Parkinson: protocolo para ensaio clínico randomizado**. [S.l.]: Brasil, 2018.

NIETZSCHE, Friedrich Wilhelm.; HOLLINGDALE, R. J. .. **Twilight of the idols and, the Anti-Christ**. [S.l.]: Penguin Books, 1990.

OH, Seung-Hyeon; LEE, Yu-Ri; KIM, Hyoung-Nam. A Novel EEG Feature Extraction Method Using Hjorth Parameter. **International Journal of Electronics and Electrical Engineering**, p. 106–110, 2014.

OIKONOMOU, V. P. *et al.* A Comparison Study on EEG Signal Processing Techniques Using Motor Imagery EEG Data. *In: IEEE*, jun. 2017.

OPPENHEIM, A. V.; SCHAFFER, R. W. **Discrete-time signal processing**. [S.l.]: Pearson, 2010.

OPPENHEIM, A. V.; WILLSKY, A. S.; NAWAB, S. H. **Signals & systems**. [S.l.]: Prentice Hall, 1997.

RASHED-AL-MAHFUZ, Md *et al.* Artifact suppression and analysis of brain activities with electroencephalography signals. **Neural regeneration research**, v. 8, n. 16, p. 1500–13, 5 jun. 2013.

ROSSO, O. A. *et al.* Wavelet entropy: a new tool for analysis of short duration brain electrical signals. **Journal of Neuroscience Methods**, v. 105, n. 1, p. 65–75, 30 jan. 2001.

RUSSELL, S. J. *et al.* **Artificial intelligence: a modern approach**. [S.l.]: Pearson, 2021.

SCHAUMANN, R.; XIAO, H.; VAN VALKENBURG, M. E. **Analog filter design**. [S.l.]: Oxford University Press, 2011.

SHARMA, R.; PACHORI, R.; ACHARYA, U. An Integrated Index for the Identification of Focal Electroencephalogram Signals Using Discrete Wavelet Transform and Entropy Measures. **Entropy**, v. 17, n. 8, p. 5218–5240, 27 jul. 2015.

SILVA, L. C. P. D. **Análise de sinais de EEG após reabilitação com imagética motora em indivíduos pós-AVC**. [S.l.]: Universidade Federal do Rio de Janeiro, jul. 2018.

SOUTO, D. O. **Imagética motora: desenvolvimento, neuropsicologia e uso na reabilitação de crianças com paralisia cerebral hemiplégica**. [S.l.: S.n.].

SOUZA, L. K. M. de *et al.* A imagética motora como estratégia para a reabilitação neurológica pós acidente vascular encefálico (AVE). *In: Dinamismo e Clareza no Planejamento em Ciências da Saúde 4*. [S.l.]: Atena Editora, 2021. p. 37–41.

STECKLOW, M. V.; INFANTOSI, Antonio Fernando Catelli; CAGY, Maurício. Alterações na banda alfa do eletrencefalograma durante imagética motora visual e cinestésica. **Arquivos de Neuro-Psiquiatria**, v. 65, n. 4a, p. 1084–1088, dez. 2007.

SUBASI, Abdulhamit; ERÇELEBI, Ergun. Classification of EEG signals using neural network and logistic regression. **Computer Methods and Programs in Biomedicine**, v. 78, n. 2, p. 87–99, 1 maio 2005.

SUKHORUKOVA, N.; KELLY, L. k-Means Clustering in EEG (Brain Waves) Timeseries. *In: [S.l.: S.n.]*. p. 625–631.

VAN DER LUBBE, Rob H. J. *et al.* Frontal brain areas are more involved during motor imagery than during motor execution/preparation of a response sequence. **International Journal of Psychophysiology**, v. 164, p. 71–86, jun. 2021.

VAZ, Y. **Extração de características para a classificação de imagética motora em interfaces cérebro-computador**. São Carlos: Universidade de São Paulo, 9 nov. 2016.

VENKATACHALAM, K. *et al.* A novel method of motor imagery classification using eeg signal. **Artificial Intelligence in Medicine**, v. 103, p. 101787, 1 mar. 2020.

YEUNG, K. Y.; RUZZO, W. L. Details of the adjusted rand index and clustering algorithms, supplement to the paper an empirical study on principal component analysis for clustering gene expression data. **Bioinformatics**, v. 17, n. 9, p. 763–774, 2001.

YUAN, Han *et al.* Negative covariation between task-related responses in alpha/beta-band activity and BOLD in human sensorimotor cortex: An EEG and fMRI study of motor imagery and movements. **NeuroImage**, v. 49, n. 3, p. 2596–2606, fev. 2010.

ZHAO, Yuyang *et al.* Effects of visually induced motor imagery-based brain-computer interface training on motor function in patients with incomplete spinal cord injury: a small-sample exploratory trial. **Frontiers in Neurology**, v. 17, 3 fev. 2026.

ZHU, Tingting *et al.* A study of motor imagery EEG classification based on feature fusion and attentional mechanisms. **Frontiers in Human Neuroscience**, v. 19, 16 jul. 2025.